

Ramirez, Delia

From: Wesner-Early, Caryn (ASRC)
Sent: Tuesday, February 12, 2002 11:53 AM
To: Ramirez, Delia
Subject: FW: dates for 2 GenBank entries (allen) (UPD#43785)

Examiner Ramirez -

Here are your publication dates - let me know if you need anything else.

Caryn
Caryn S. Wesner-Early, MSLS
Technical Information Specialist
Biotechnology and Chemical Library
U.S. Patent and Trademark Office
Phone: (703) 308-4501
Fax: (703) 308-4496
caryn.wesner@uspto.gov

-----Original Message-----

From: update@ebi.ac.uk [mailto:update@ebi.ac.uk]
Sent: Tuesday, February 12, 2002 9:50 AM
To: Caryn.Wesner@uspto.gov
Subject: Re: dates for 2 GenBank entries (allen) (UPD#43785)

Dear colleague

You have requested release date information for EMBL entry(s):Z49617 and Z49260.

The date given in the first DT line of this entry is the date the entry first became available for public disclosure.

Example:

DT dd-month-year (Rel. nn, Created)

The date in the Submission Reference RL line

Example:

RL Submitted (dd-month-year) to the EMBL/GenBank/DDBJ databases.
is the date that EMBL received the submission from the submitter.

DISCLAIMER: PLEASE NOTE

No guarantee is given that release date given in the entry is accurate. The actual release date will depend on the availability of network services.

No guarantee is given that the entry is complete and accurate. In particular the sequence data of the entry may not conform to the sequence data of the original publication where the sequence was first disclosed to the public.

I enclose a copy of the entry(s).

Yours sincerely

Yvonne Allen

ID SCYJR117W standard; DNA; FUN; 1825 BP.
XX
AC Z49617; Y13136;
XX

THIS PAGE BLANK (0.000)

```

SV      Z49617.1
XX
DT      06-OCT-1995 (Rel. 45, Created)
DT      11-AUG-1997 (Rel. 52, Last updated, Version 2)
XX
DE      S.cerevisiae chromosome X reading frame ORF YJR117w
XX
KW
XX
OS      Saccharomyces cerevisiae (baker's yeast)
OC      Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC      Saccharomycetales; Saccharomycetaceae; Saccharomyces.
XX
RN      [1]
RP      1-1825
RA      Rose M., Koetter P., Entian K.D.;
RT      ;
RL      Unpublished.
XX
RN      [2]
RP      1-1825
RA      MIPS;
RT      ;
RL      Submitted (25-SEP-1995) to the EMBL/GenBank/DDBJ databases.
RL      Data collected by MIPS on behalf of the European yeast chromosome X
RL      sequencing project. MIPS at the Max-Planck-Institut fuer Biochemie, Am
RL      Klopferspitz 18a D-82152 Martinsried, FRG; E-mail: Mewes@mips.embnnet.org
XX
DR      SGD; S0003878; YJR117W.
DR      SWISS-PROT; P47154; ST24_YEAST.
XX
FH      Key                Location/Qualifiers
FH
FT      CDS                343..1704
FT                        /db_xref="SGD:S0003878"
FT                        /db_xref="SWISS-PROT:P47154"
FT                        /note="ORF YJR117w"
FT                        /gene="STE24"
FT                        /protein_id="CAA89647.1"
FT                        /translation="MFDLKTILDHPNIPWKLIISGFSIAQFSFESYLTYRQYQKLSETK
FT      LPPVLEDEIDDETfHKSRNYSRAKAKFSIFGDVYNLAQKLVFIFYDLFPKIWHMAVSLL
FT      NAVLPVRFHMVSTVAQSLCFLGLLSSLSTLVDLPLSYSHFVLEEKFGFNKLTVQLWIT
FT      DMIKSLTLAYAIGGPILYLFLKIFDKFPTDFLWYIMVFLFVVQILAMTIIPVFIMPMFN
FT      KFTPLEDGELKKSIESLADRVGFPLDKIFVIDGSKRSSHSNAYFTGLPFTSKRIVLFD
FT      LVNSNSTDEITAVLAHEIGHWQKNHIVNMVIFSQ LHTFLIFSLFTSIYRNTSFYNTFGF
FT      FLEKSTGSFVDPVITKEFPPIIIGFMLFNDLLTPLECAMQFVMSLISRTHEYQADAYAKK
FT      LGYKQNLCLRALIDLQIKNLSTMNVDPLYSSYHYSHPTLAERLTALDYVSEKKKN"
FT      source              1..1825
FT                        /chromosome="X"
FT                        /db_xref="taxon:4932"
FT                        /organism="Saccharomyces cerevisiae"
XX
SQ      Sequence 1825 BP; 570 A; 397 C; 280 G; 578 T; 0 other;
acctaccttt ttttctatct tcaacaacga aacgccttac acacacacac acatacatct      60
acatacatat atacaaatat acatatatgt aaacttgtat attcattcct attaaccaaa      120
aagaggcaat taaacttttc cctctttttc tacgtcattt actcaaaaac tctaattcct      180
tcgtctctgt tctgccattt tctccagaaa aaaatcgacg ggaaataaaa aaaaaaagac      240
aacgaacaag agaaaaagtt cggaattat aaaccacttc tataattaac aggaaaagga      300
aggaaaaaaa aggaggaaat agaaaactgc aggcctttat tcatgtttga tcttaagacg      360
attctcgacc atcctaatat cccgtggaaa ttaatcattt ctgggttctc gattgcccac      420
ttttctttcg aatcttactt gacgtacaga cagtaaccaga agctatctga aacaaagttg      480
ccacctgtgc tggaagacga aattgatgat gaaacttttc ataaatcaag gaactactcc      540
cgggccaaag ccaagttctc cattttcggt gacgtctata acctagccca aaagctagtt      600
ttcatcaaat acgacctctt ccctaaaate tggcacatgg ccgtttcttt attgaatgca      660
gtcctgccag tcagatttca tatggtctcc actgtcgcac agagtttatg cttcttggtg      720
ctcttatcca gtttgtctac cttggttgat ttgccactct cttactatag ccattttgtc      780
ctggaagaaa aatttggttt caataaattg accgtccaac tatggatcac cgatatgatc      840

```

THIS PAGE BLANK (0.370)


```

aagagtctga ctttggcgta tgctattggt ggcccaatcc tttacctggt ccttaagatc      900
tttgataaat tccctactga tttcctttgg tacattatgg tcttcttggt cgttgtccaa      960
atccttagcca tgacaatcat tccagtcttc atcatgccca tgtttaataa gttcactcca     1020
ttggaggacg gtgaactgaa aaaatctatt gaaagtttgg ccgatagagt tgggttccct     1080
ctagataaga tttttgtcat tgacggctca aaaagatctt ctcatcmeta cgcataatttc     1140
acaggtttgc cattcacctc caagagaatt gttttgttcg acactttagt gaacagtaat     1200
tctactgatg aaattacggc tgttttggcc catgaaatcg gtcactggca aaaaaaccac     1260
atcgtttaata tggatcatctt tagtcaattg cacaccttcc tcattttctc ccttttcacc     1320
agcatctaca gaaatacatc attttacaac accttcgggt ttttcttaga gaagtccact     1380
ggcagttttg ttgatcccggt tatcactaag gaattcccca ttatcattgg atttatgtta     1440
tttaacgact tattaactcc actcgaatgt gccatgcaat tcgtgatgag tttaatttcc     1500
agaactcatg aatatcaagc tgatgcttat gctaaaaaat tgggctacaa gcaaaatcta     1560
tgtagggtc taattgatct acaaatcaaa aacctttcca ccatgaatgt agatcctctg     1620
tattctagct atcattattc ccatccaact ctagctgaaa gattgaccgc tctagactat     1680
gttagtgaaa agaagaaaaa ctaatctata gagtacacat attagcatgt accgttaaat     1740
tcagcttcgt tatgtctata tctacataca tacacaggta tctactataa gaataaagga     1800
aagaaaaaat aaacgattaa acatt                                           1825

```

//

```

ID      SC8156      standard; DNA; FUN; 29987 BP.
XX
AC      Z49260; Z71257;
XX
SV      Z49260.1
XX
DT      16-MAY-1995 (Rel. 43, Created)
DT      11-AUG-1997 (Rel. 52, Last updated, Version 1)
XX
DE      S.cerevisiae chromosome XIII cosmid 8156.
XX
KW      BUL1; cytochrome b5; DAG1; initiation factor 1A; inorganic pyrophosphatase;
KW      IPP2; nitrate reductase; orotate phosphoribosyltransferase; tau element;
KW      TIF11; TPS3; transfer RNA-Gln; trehalose-phosphate synthase; U6 snRNP;
KW      URA10.
XX
OS      Saccharomyces cerevisiae (baker's yeast)
OC      Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC      Saccharomycetales; Saccharomycetaceae; Saccharomyces.
XX
RN      [1]
RP      1-29987
RA      Lye G., Churcher C.M.;
RT      ;
RL      Unpublished.
XX
RN      [2]
RP      1-29987
RA      Barrell B., Rajandream M.A.;
RT      ;
RL      Submitted (12-MAY-1995) to the EMBL/GenBank/DDBJ databases.
RL      Saccharomyces cerevisiae chromosome XIII sequencing project, Sanger Centre,
RL      Hinxton Hall, Hinxton, Cambridge CB10 1RQ E-mail: barrell@sanger.ac.uk
XX
DR      SGD; S0004873; YMR260C.
DR      SGD; S0004874; YMR261C.
DR      SGD; S0004875; YMR262W.
DR      SGD; S0004876; YMR263W.
DR      SGD; S0004877; YMR264W.
DR      SGD; S0004878; YMR265C.
DR      SGD; S0004879; YMR266W.
DR      SGD; S0004880; YMR267W.
DR      SGD; S0004881; YMR268C.
DR      SGD; S0004882; YMR269W.
DR      SGD; S0004883; YMR270C.
DR      SGD; S0004884; YMR271C.
DR      SGD; S0004885; YMR272C.

```

THIS PAGE BLANK (b)(7)(D)

```

DR   SGD; S0004886; YMR273C.
DR   SGD; S0004887; YMR274C.
DR   SGD; S0004888; YMR275C.
DR   SWISS-PROT; P28239; IPY2_YEAST.
DR   SWISS-PROT; P30402; PYRX_YEAST.
DR   SWISS-PROT; P38426; TPS3_YEAST.
DR   SWISS-PROT; P38428; YM8E_YEAST.
DR   SWISS-PROT; P38429; YM8D_YEAST.
DR   SWISS-PROT; P38430; YM8C_YEAST.
DR   SWISS-PROT; P38912; IF1A_YEAST.
DR   SWISS-PROT; P48524; BUL1_YEAST.
DR   SWISS-PROT; P49960; PR24_YEAST.
DR   SWISS-PROT; P50111; ZDS1_YEAST.
DR   SWISS-PROT; P53437; RRN9_YEAST.
DR   SWISS-PROT; Q03496; YM8B_YEAST.
DR   SWISS-PROT; Q03508; YM8F_YEAST.
DR   SWISS-PROT; Q03516; YM8G_YEAST.
DR   SWISS-PROT; Q03525; YM8H_YEAST.
DR   SWISS-PROT; Q03529; SCS7_YEAST.
DR   SWISS-PROT; Q03530; RCE1_YEAST.

```

XX

CC Notes:

XX

CC All CDS over 100 codons have been analysed. CDS that are
 CC completely overlapped and those that are overlapped by more than
 CC 50%
 CC of their length by a larger CDS have been omitted from this
 CC analysis.
 CC Details of the omitted CDS are available on request.
 CC The more significant matches with motifs in the PROSITE database
 CC are
 CC also included but some of these may be fortuitous.
 CC The length in codons and the calculated codon adaptation index
 CC (CAI)
 CC is given for each CDS.

XX

CC Cosmid 8156 is overlapped at the start by cosmid 9920, embl entry
 CC SC9920,
 CC accession no. Z48639 and at the end by cosmid 8021.

XX

| Key | Location/Qualifiers |
|--------------|---|
| source | 1..29987 /chromosome="XIII" /db_xref="taxon:4932" /organism="Saccharomyces cerevisiae" /strain="AB972" /clone="cosmid 8156" /map="13R" |
| misc_feature | 1..140 /note="overlap with SC9920 Z48639 S.cerevisiae chromosome XIII cosmid 9920" |
| CDS | complement(<1..356) /db_xref="SWISS-PROT:Q03496" /note="YM8156.01c, unknown, len: 118, CAI: 0.14, overlaps and extends YM9920.13c" /product="unknown" /protein_id="CAA89242.1" /translation="MTTDVQFSSQEIELFRVKEFLIANNPAKINNENKDAVLTQIEHDF RYLIQYIKDGLPNLNSTRILFPDTFSICLLRSHQIIASKKIDSQEFLSAVKEQLLTEA NANIIFEYVLDFWA" |
| CDS | complement(851..1312) /db_xref="SGD:S0004873" /db_xref="SWISS-PROT:P38912" /note="YM8156.02c TIF11 gene, IF1A_YEAST, P3891, initiation factor 1a" /gene="TIF11" |

THIS PAGE BLANK (L.3PT0)

```

FT      /product="Tif11p"
FT      /protein_id="CAA89243.1"
FT      /translation="MGKKNTKGGKKGRRGKNDSDGPKRELIYKEEGQEYAQITKMLGNG
FT      RVEASCFDGNKMAHIRGKLRKKVWVGQDIIILVSLRDFQDDQCDVVHKYNLDEARTLK
FT      NQGELPENAKINETDNFGFESDEDVNFEGNADEDEDEGEDEELDIDDI"
FT      CDS      complement(1678..4842)
FT      /db_xref="SGD:S0004874"
FT      /db_xref="SWISS-PROT:P38426"
FT      /note="YM8156.03c, TPS3 gene, len: 1054, CAI: 0.21,
FT      SW:TPS3_YEAST P38426; alpha,alpha-trehalose-phosphate
FT      synthase and PIR:S27471; PS00213 Lipocalin signature;
FT      translation in TPS3_YEAST starts at amino acid 33 in this
FT      sequence"
FT      /gene="TPS3"
FT      /product="Tps3p"
FT      /protein_id="CAA89244.1"
FT      /translation="MTIIVASLFLPYTPQFEADVNTSDTAKLVNESSMIKVCNNQELSN
FT      NKQERSSSVTSASSHYIGLPQEAQINGEPLQRANVGSPATGVNYHNEMEMLSSEQFLEE
FT      LTANATHAANS GIPPANPNVSSGSTAQRPSVEEFFSAPSARVCSPSQEASASSISASRS
FT      SAHNDLSSSLMKPNLSFDSHPVRSSSKSAVITPVSKSPDVPDPAVVDVAKVREEF
FT      QQQASLPSMKRVSGSTAGDSSIASSSSNLRYSQQFQDNFIEDTDSDDIDSLETDATK
FT      KYNVPKFGGYSNNAKLRLASLMRNSYELFKHLPWTIVDSDKGNGLSKNAVNIAVAEKT VK
FT      EPVSWVGTMGIPTDELPHVCHKISKKLEQDFSSFPVVTDDITFKGAYKKNYAKQILWPT
FT      LHYQIPDNPNSKAFEDHSWDYYQKVNQKFSRIVSVYKPGDTIWIHDYHMLVPMQVRE
FT      KLPKAKIGFPLHVSFPSEVFRCLANRERILEGIIGANFVGFGQTKEYKRHFLQTCNRL
FT      AADVSNDEVKYHCNIVSVMYAPIGIDYYHLTSQLRNGSVLEWRQLIKERWRNKKLIVCR
FT      DQFDRIRGLQKKMLAYERFLIENPEYIEKVVLIIQICIGKSSDPEYERQIMVVDRINSL
FT      SSNISISQPVVFLHQDLDFQAQYLALNCEADVFLVDALREGMNLTCHEFIVSSFENAPL
FT      LLSEFTGSSSVLKEGAILINPWDINHVAQSIKRSLEMSPEEKRRRWKKLFKSVIEHDS
FT      NWITKCFEYINNAWESNQETSTVFNLAPKFCADYKASKKHLFIFKISEPPTSRMLSLL
FT      SELSSNNIVYVLSSFTKNTFESLYNGVLNIGLIAENGAYVRVNGSWYNIVEELDWMKEV
FT      AKIFDEKVERLPGSYKYIADSMIRFHTENADDQDRVPTVIGEATHINTLFDDRD IHAY
FT      VHKDIVFVQQTGLALAAAEFLMKFYNSGVSPDTSRISLRTSSSMVGNKKHFFQNVQV
FT      DFVCSVGSTSPIIEPLFKLVKQVEKNNLKFYGTILYGSRRSTYAKEHINGVNELFTIL
FT      HDLTAA"
FT      misc_feature      complement(2260..2301)
FT      /note="PS00213 Lipocalin signature"
FT      CDS      5199..6140
FT      /db_xref="SGD:S0004875"
FT      /db_xref="SWISS-PROT:P38430"
FT      /note="YM8156.04, unknown, len: 313, CAI: 0.11, previously
FT      noted ORF, SW:YTP3_YEAST P38430"
FT      /product="unknown"
FT      /protein_id="CAA89245.1"
FT      /translation="MNKLVDACHVITDPDNTFCGDDGGSQGTLCVMSSNPYDWNNLK
FT      KLAGRSTSKNDICVGFVHPWYSHLFYVGSRDKVSHYQDVLEYKNEEQFDSLQVLPPE
FT      PLDLEEYIKREFNDTLVSVIGEIGLDKFLRLPANGFYMONEKARLT TVKVKLSHQETVF
FT      RRFWRRLARHTSKPISIHDKCHGKLN DICNEELLTYHSVKICLHSYTGSKETLLGQWLK
FT      KFPPDRIFVLSKWINFKDPEEGDALVRSLPSTCILTETDYPIDNPDPYQKALTEQLQ
FT      YLNAQIARAWDETLDASQAALRVYENFQKFIK"
FT      CDS      6392..6997
FT      /db_xref="SGD:S0004876"
FT      /db_xref="SWISS-PROT:P38429"
FT      /note="YM8156.05, unknown, len: 301, CAI: 0.12,
FT      SW:YTP2_YEAST P38429 previously noted ORF"
FT      /product="unknown"
FT      /protein_id="CAA89246.1"
FT      /translation="MARPVNTNAETESRGRPTQGGGYASNNNGSCNNNNNGSNNNNNNNN
FT      NNNNNNSNNNNNGPTSSGRNTGKQRLTAAQQQYIKNLIETHITDNHPDLRPKSHPMDF
FT      EEYTD AFLRRYKDHFLQDVPDNLTLQGYLLGSKLGAKTYSYKRNTQGQHDKRIHKRDLA
FT      NVVRRHFDEHSIKETDCIPQFIYKVNQKKKFKMEFRG"
FT      CDS      7278..7889
FT      /db_xref="SGD:S0004877"
FT      /db_xref="SWISS-PROT:P38428"
FT      /note="YM8156.06, unknown, len: 301, CAI: 0.13,
FT      SW:YTP1_YEAST P38428, extends previously noted ORF"
FT      /product="unknown"

```

THIS PAGE BLANK (3PT0)

```

FT      /protein_id="CAA89247.1"
FT      /translation="MEDSRLLITLILVFGVIFLKKFFQSNQHPSAQRLSATGVNAHGRP
FT      QGSTQNALRRRTGRVNGGHPVTTQMVETVQNLAPNLHPEQIRYSLENTGSVEETVERYLR
FT      GDEFSFPFGFEPSPRPMGANAAVDNNAAGGGEFNDPRKKNMICAENLLDKFHVLDLNE
FT      SNLSFKDLIDIEERKRLVWQARKNLETKLQSDKDLQSLLT"
FT      CDS      complement(8013..9398)
FT      /db_xref="SGD:S0004878"
FT      /db_xref="SWISS-PROT:Q03508"
FT      /note="YM8156.07c, unknown, len: 461, CAI: 0.11"
FT      /product="unknown"
FT      /protein_id="CAA89248.1"
FT      /translation="MEEFEEFRKKGEMSSRCGNHRVLRKWNSCACELAVPFEVPEHAIT
FT      KLHIYDFDNTLFATPGPTEQLYTRELLNLLTSSTLPNGGWNEPGFLQAAIEISKTKPR
FT      RYSWNADIVKLAEEYSAKDTISIVLTGREESKFHKLIEHALQTARSHWKCSENEFRFN
FT      AVCLKKRAISEYTSKYKKELMRDFLEYYPSSLRELSIYDDRIHQIDAFKSFHSLDLPLR
FT      KWSAIPVRPFTKSLPREQELEMVMDMVRKNNSQALSTSQKFDLRRTPRQIGYILCTASH
FT      RLLSIEVIKYLKRRKGRRTFRPKLYEHPYIPCAEPGKDIPALEIAKVWSNNDTRTFDS
FT      EKKVQHISQIFYLEQPGKCIVHFQVTDLAVIASAHNRRKPLEVYFKATPEPNRYTFTL
FT      FPYIVTGHFYKRDRIEDLEVVTTERLINCKEDIHWVPLDNTIPIKAFFGRFAKLAAIPC
FT      SNA"
FT      CDS      9991..12852
FT      /db_xref="SGD:S0004879"
FT      /db_xref="SWISS-PROT:Q03516"
FT      /note="YM8156.08, unknown, len: 953, CAI: 0.20, probable
FT      integral membrane glycoprotein"
FT      /product="unknown"
FT      /protein_id="CAA89249.1"
FT      /translation="MNSTNSTNSTTTATSTNTSTQQVVTSLVSNGTIFGVFVIAFLILR
FT      IKLKRIYEPKSSFNLINEEKKPEPLPQGVWQWLKPLLKKSDFVVIQQAGLDGYFFLRYL
FT      FIITAIYCAVMSYIFPILLSINASNGNHESGLNQLAYQNVKHRGRYFAHVFCGWIFFWG
FT      FLYIIYRELYFYTSMKQAVLASPRYAKKLSSRTVLFQTVPKQYLSEEEFSKLFQDGVKRV
FT      WIARGSGSIEAMVKARDNMAIQLEGAETKYLKAALKKIKKLNKKSPQLSVSDNIAEYVP
FT      DKKRPHHKINKVAKFFFGKKVDTISYIKEELPKLNQKVKALQEDHENSFPNSVFVEFE
FT      SQYQAQVAAQITTYHAPLFMTPVYIGIEPSDVVWFNLRMFWVERLGREVSAVSAIVALV
FT      ILWAFPVAFVGMISNITSLTNEVKWLKFIYKLPKQLLGLLTSAPTVALAVLMSFLPKF
FT      IRGMAITQGAPSKQVVEYFTQAYFAFQVIQVFLVTTLSAATSTVTEIVKEPTKAMD
FT      LASNLPKASNPFMSYVILQGLSISGALLQIVPLILFYVLGAFLDGTVRKKWNRFGLS
FT      SMQWGTAFFPVYTNLAVITFSYSIISPLILLFAAVAFFLLYIAYLYNLTYVYQESPDARG
FT      IYYPRALFQITVIGIYIGQICLLGLFAVGKGGPIVLQVIGICVTVLIHLHLSAAPDHL
FT      KVIPVDTMKPLDGVSDTPSFKNYKGIESTKVKKNTFGANIDMDGKELPEFPKIKYHK
FT      RSESVTEQQVENSIFSENTFEYQFNPAEANADGHAINAENLIEDVPLLADGDTMKIPP
FT      APWWKRFLKPHIYYSYKAVKSRLEIYGLVDPDERVNDFDISHAYDPAVSAQCPDWI
FT      PRDPFGFSKLLISDVSGVVMNDENATIDENLKFTLRDVPPPYNVDVKDEANGEANG
FT      TASKENNPFPADPKYKEESRSV"
FT      CDS      13245..14177
FT      /db_xref="SGD:S0004880"
FT      /db_xref="SWISS-PROT:P28239"
FT      /note="YM8156.09, IPP2 gene (PPA2), len: 310, CAI: 0.14,
FT      SW:IPY2_YEAST, P28239, inorganic pyrophosphatase,
FT      mitochondrial; contains PS00387 Inorganic pyrophosphatase
FT      signature"
FT      /gene="IPP2"
FT      /product="Ipp2p"
FT      /protein_id="CAA89250.1"
FT      /translation="MNLRLMNAITSKARSIERLQTLNLSIRNHRQFSTIQQGSKYTL
FT      GFKKYLTLNGEVGSFFHDVPLDLNEHEKTVNMIVEVPRWTTGKFEISKELRFNPVQD
FT      TKNGKLRFVNIPYHGYIHNYGAIPQTWEDPTIEHKLKCDVALKGDNDPLDCCEIGS
FT      DVLEMGSIKKVKVLGSLALIDDGELDWKVIVIDVNDPLSSKIDDLKIEEYFPGILD
FT      REWFRKYKVPAGKPLNSFAFHEQYQNSNKTIQTIKKCHNSWKNLISGSLQEKYDNLN
FT      ERAGNGVTLEDVSVKPPSQIPPEVQKWWYV"
FT      misc_feature      13698..13718
FT      /note="PS00387 Inorganic pyrophosphatase signature"
FT      CDS      complement(14361..15695)
FT      /db_xref="SGD:S0004881"
FT      /db_xref="SWISS-PROT:P49960"
FT      /note="YP8156.10c, U6 snRNP gene, len: 444, CAI: 0.12,
FT      amino acids 200-288 identical to PIR:A39562 A39562 U6 snRNP

```

THIS PAGE BLANK (USPTO)


```

FT      protein PRP24"
FT      /gene="U6sNRP"
FT      /product="U6sNRPp"
FT      /protein_id="CAA89251.1"
FT      /translation="MEYGHHPDPSKRPLDEGSPAAAGLTSKKANEALTRNRELT'TVLV
FT      KNLPKSYNQNKVYKYFKHCGPIIHVDVADSLKKNFRFARIEFARYDGALAAITKTHKVV
FT      GQNEIIVSHLTCTLWMTNFPSPYQNRNIRDLLQDINVVALSIRLPSLRFNTSRRFAYI
FT      DVTSKEDARYCVEKLNGLKIEGYTLVTKVSNPLEKSKRTDSATLEGREIMIRNLSTELL
FT      DENLLRESFEGFGSIEKINIPAGQKEHSFNNCCAFMVFENKDSAERALQMNRSLLGNRE
FT      ISVSLADKKPFLENERVKRLLASRNSKELETLICLFPLSDKVSPSLICQFLQEEIHINE
FT      KDIRKILLVSDFNCAIIIFRDSKFAAKMLMILNGSQFQGVIRSGTINDMKRYNNQQN
FT      HSMKHVKPSCINMMEKGNLQVKKKIPDKQEOMSNDDFRKMFLGE"
FT      CDS      16137..16565
FT      /db_xref="SGD:S0004882"
FT      /db_xref="SWISS-PROT:Q03525"
FT      /note="YM8156.11, unknown, len: 142, CAI: 0.13, Lys-, Ser-
FT      rich hypothetical protein"
FT      /product="unknown"
FT      /protein_id="CAA89252.1"
FT      /translation="MAVLNLLQNEAVATAVSKSSSPLYRWFVKGEGLKGTITNLGKKEE
FT      ASFVVSASSSSKGKKRRRDEDDNKVKRKKLKKDKKTSNDSESKKKKKKSKKESKKGK
FT      KSKHSSDEGDKSKHKKSKSKSKHKKEESSARRDRKEHI"
FT      CDS      complement(16800..17897)
FT      /db_xref="SGD:S0004883"
FT      /db_xref="SWISS-PROT:P53437"
FT      /note="YM8156.12c, unknown, len: 365, CAI: 0.14"
FT      /product="unknown"
FT      /protein_id="CAA89253.1"
FT      /translation="MSDLDEESQIETQIDAPIEDIIRGSELTTTADKETLKSANELLD
FT      SLEHSHRVDLSLHLYSAYLLKRLLYKANEEKHFYEVNQFVKQIKDNWTSWPNPNTIID
FT      PSVDKLYEDIPEGIANVSVQGEISNRALMHASDMMRVELDAQWQKFLSKSALDHDVTL
FT      DVDELNIPNEISRNILVKLDSLFEGLHDKIAKENEFQVQDKHSNNIRANQIDDEPMQA
FT      NRRRIKYTYHDLVSRGCEMNEDMTDIYMKSLLELYNDIPEKYKKRKFRLPKQILKKYHQPK
FT      KTSSYLKELLSKTREDFIPVEKLLKDKRLTSKDKSKLQRLNREETEDALNKRTPFFQVKG
FT      YLEDENEISDYELDDCLIELPNGNI"
FT      CDS      complement(18339..19022)
FT      /db_xref="SGD:S0004884"
FT      /db_xref="SWISS-PROT:P30402"
FT      /note="YM8156.13c, URA10, len: 227, CAI: 0.14,
FT      SW:PYRX_YEAST P30402, contains orotate
FT      phosphoribosyltransferase 2; PS00103 Purine/pyrimidine
FT      phosphoribosyl transferases signature"
FT      /gene="URA10"
FT      /product="Ura10p"
FT      /protein_id="CAA89254.1"
FT      /translation="MSASTTSLEEYQKTFLELGLGLECKALRFGSFKLNSGRQSPYFFNLS
FT      LFNSGKLLANLATAYATAIIQSELKFDVIFGPAYKGIPLAAIVCVKLAIEIGGTFQGIQ
FT      YAFNRKKVKDHGEGGIIVGASLEDKRVLIIDVMTAGTAINEAFETISIAQGRVVGCV
FT      ALDRQEVIHESDPERTSATQSVSKRYNVPVLSIVSLTQVVQFMGNRLSPEQKSAIENYR
FT      KAYGI"
FT      misc_feature complement(18594..18632)
FT      /note="PS00103 Purine/pyrimidine phosphoribosyl
FT      transferases signature"
FT      tRNA      complement(19720..19791)
FT      /note="tRNA-Gln gene, anticodon ctg, len: 72"
FT      /gene="tRNA-Gln"
FT      LTR      complement(20011..20380)
FT      /note="solo tau element"
FT      CDS      complement(21097..22251)
FT      /db_xref="SGD:S0004885"
FT      /db_xref="SWISS-PROT:Q03529"
FT      /note="YM8156.14c, unknown, len: 384, CAI: 0.25, similar to
FT      cytochrome b5/nitrate reductase family; PS00191 Cytochrome
FT      b5 family, heme-binding domain signature"
FT      /product="unknown"
FT      /protein_id="CAA89255.1"
FT      /translation="MSTNTSKTLELFSKKTVQEHNTANDCWVITYQNRKIYDVTRFLSEH

```

THIS PAGE BLANK (0.3870)

```

FT      PGGDESILDYAGKDITEIMKSDSDVHEHSDSAYEILEDEYLGIGLATDEEAARLLTNKNH
FT      KVEVQLSADGTEFDSTTFVKELPAEEKLSIATDYSNDYKHKHFLDLNRPLLMQILRSDF
FT      KKDFYVDQIHRPRHYGKGSAPLFGNFLEPLTKTAWVVPVAVLPPVVVYHMGVALKNMNQ
FT      LFACFLFCVGVFVWTLIEYGLHRFLFHFDLWLPESNIAFATHFLLHGCHHYLPMDKYRL
FT      VMPPTLFLVILCAPFYKLVFALLPLYWAYAGFAGGLFGYVCYDECHFFLHHSKLPPFMRK
FT      LKKYHLEHHYKNYQLGFGVTSWFWDDEVFGTYLGPDAPLSKMKYE"
FT      misc_feature complement(22108..22131)
FT      /note="PS00191 Cytochrome b5 family, heme-binding domain
FT      signature"
FT      CDS complement(22707..25454)
FT      /db_xref="SGD:S0004886"
FT      /db_xref="SWISS-PROT:P50111"
FT      /note="YM8156.15c, unknown, len: 915, CAI: 0.14"
FT      /product="unknown"
FT      /protein_id="CAA89256.1"
FT      /translation="MSNRDNESMLRTTSSDKAIASQRDKRKSEVLIAAQSLDNEIRSVK
FT      NLKRLSIGSMDLLIDPELDIKFGGESSGRRSWGTTSSSASMPSDTTTNNTRYSDPTP
FT      LENLHGRGNSGIESSNKTQGNLYLGKKGVHSPSRKLNANVLKKNLLWVPANQHPNVKP
FT      DNFLVLVQDTLQNIQLSDNGEDNDGNSNENNDIEDNGEDKESQSYENKENNTINLNRGL
FT      SRHGNASLIRRPSTLRRSYTEFDDNEDDDNKGDSASETVNKVEERISKIKERPVSRLDI
FT      TEELTKISNSAGLTDNDAITLARTLSMAGSYSDKKDQPQPEGHYDEGDIGFSTSQANTL
FT      DDGEFASNMPINNMTWPERSSLRSSRFNTYRIRSQEKEVEQSVDEMKNDDERLKL
FT      TKNTIKVEIDPHKSPFRQQDESDENMSSPGSIGDFQDIYNHYRQSSGEWEQEMGIEKEA
FT      EEVPVKVRNDTEVQDLELREGTTDMVKPSATDDNKETKRHRRRNGWTWLNKMSREDDN
FT      EENQDDENEENEVDQSQRMELDNSKKHYISLFNGGEKTEVSNKEEMNSSTSTATSQTRQ
FT      KIEKTFANLFRRKPHHKHDASSPSSSSPSSSPSIPNNDVHVRVRKSKKLGNKSGREPV
FT      EPIVLRNRPRPHRHHHSRHGSQKISVKTLKDSQPQQQIPLQPQLEGAIEIEKKEESDSE
FT      SLPQLQPAVSVSSTKSNSRDREEEAAKKKNKRSNTTEISNQQHSKHVQKENTDEQKAQ
FT      LQAPAQEQVQTSVPVQASAPVQNSAPVQTSAPVEASAQTQAPAAPPLKHTSILPPRKL
FT      TADVKKPKDPNSPVQFTDSAFGFPLPLLTVSTVIMFDHRLPINVERATYRLSHLKLNS
FT      KRGLEQVLLSNFMYAYLNLVNHTLYMEQVAHDKEQQQQQQQQP"
FT      CDS complement(25838..26785)
FT      /db_xref="SGD:S0004887"
FT      /db_xref="SWISS-PROT:Q03530"
FT      /note="YM8156.16c, unknown, len: 315, CAI: 0.10"
FT      /product="unknown"
FT      /protein_id="CAA89257.1"
FT      /translation="MLQFSTFLVLLYISISYVLPLYATSQPEGSKRDNPRTIKSRMQKL
FT      TIMLISNLFVLPFLQSQLSSTTSHISFKDAFLGLGIIPGYAALPNPWQFSQFVKDLTK
FT      CVAMLLTLYCGPVLDVFLYHLLNPKSSILEDFYHEFLNIWSFRNFIFAPITEEIFYTSM
FT      LLTTYLNLIPHSYQLSQQFLWQPSLFFGLAHAAHAYEQLQEGSMTTVSILLTTCFQILY
FT      TTLFGGLTKFVFRVTGGNLWCCIIILHALCNIMFGPGPSRLNLHFTVVDKKAGRISKLV
FT      IWNKCYFALLVLGLISLKDTLQTLVGTPGYRITL"
FT      CDS complement(27125..>29987)
FT      /codon_start=2
FT      /db_xref="SGD:S0004888"
FT      /db_xref="SWISS-PROT:P48524"
FT      /note="YM8156.17c, BUL1 gene, (DAG1), len: >953, CAI: 0.14,
FT      EM_NEW:SCBUL1 D50083"
FT      /gene="BUL1"
FT      /product="Bul1p"
FT      /protein_id="CAA89258.1"
FT      /translation="DFTANSSTTMNVNANTRGRGRQKQEGGKSSRSPLHSPKSWIRS
FT      ASATGILGLRRPELAHSHSHAPSTGTGTPAGGNRSPLRRSTANATPVETGRSLTDGINNV
FT      VDVLPSEFMYNTLHRHIPQGNVDPDRHDFPPSYQEANNSTATGAAGSSADLSHQSLSTD
FT      ALGATRSSSTSNLENLIPLRTEHHSIAAHQSTAVDEDSLDIPPILDDLNDTDNIFIDKL
FT      YTLPKMSTPIBITIKTKHAPIPHVKPEEESILKEYTSGDLIHGFIITENKSQANLKEF
FT      MPYVTLSEYISIIDKVKSKRTIKRFLRMVDLSASWSYSKIALGSGVDIFPADVDYDGSV
FT      FGLNNSRVLEPGVKYKFFIFKLPQLQLLDVTCKQEHFSSHCLLPSPSGIDKYRNNCKYSG
FT      IKVNRVLGCGHLGKGPILTNDMSDDNLSINYTIDARIVGKDQKASKLYIMKEREYNL
FT      RVIPFGFDANVVGERTTMSQLNDITKLVOERLDALRKIFORLEKKEPITNRDIHGADLS
FT      GTIDDSIESDSQEIQLQRKLDQLHIKNRNNYLVNYNDLKLGHDLNDRSGNSGHNTDTSR
FT      AWGPFVESELKYKLKNKSNSSSFLNFSHFLNSSSSSSMSSSSNAGKNNHDLTGKNKERTGL
FT      ILVKAKIPKQGLPYWAPSLLRKTNVFESKSKHDQENWVRLSELIPEDVKKPLEKLDLQL
FT      TCIBSDNSLPHDPPEIQSITTELCITAKSDNSIPIKLNSELLMNKEKLTISKALYDDF
FT      HSKICEYETKFNKFLNELNLYNMNRGDRRPKELKFTDFITSQLFNDIESICNLKVS
FT      NLSNIFKKQVSTLQHSKHALSEDSISHTGNGSSSSPSSASLTPVTSSSKSSLFLPSGS

```

THIS PAGE BLANK (1-370)

FT SSTS LKFTDQIVHKWVRIAPLQYKRDINVNLEFNKDIKETLIPSFESCLCCRFYCVRVM
FT IKFENHLGVAKIDIPISVRQVTK"
FT misc_feature 29849..29987
FT /note="overlap with cosmid 8021"
XX
SQ Sequence 29987 BP; 8850 A; 5706 C; 5896 G; 9535 T; 0 other;

Sequence omitted for brevity

Updates - The EMBL Nucleotide Sequence Database
EMBL Outstation Hinxton, The European Bioinformatics Institute,
Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SD, U.K.
Tel: +44 1223 494499. Fax: +44 1223 494472

Internet electronic addresses:

datasubs@ebi.ac.uk (data submissions), update@ebi.ac.uk (updates)
datalib@ebi.ac.uk (general enquiries)
<http://www.ebi.ac.uk/> (world wide web submissions, updates, services,
info)

THIS PAGE BLANK (JPTO)

RP CHARACTERIZATION.
 RX MEDLINE-20285442; PubMed-10825201;
 RA Trublood C.E., Boyartchuk V.L., Pictologlou E.A., Rozema D.,
 RA Poulter C.D., Rine J.;
 RT "The Caax proteases, Aicp and Rcelp, have overlapping but distinct
 RT substrate specificities.";
 RT Mol. Cell. Biol. 20:4381-4392(2000).
 RN (8)
 RP SUBCELLULAR LOCATION.
 RX MEDLINE-98409630; PubMed-9736709;
 RA Schmidt W.K., Tam A., Fujimura-Kamada K., Michaelis S.;
 RT "Endoplasmic reticulum membrane localization of Rcelp and Ste24p.
 RT Yeast proteases involved in carboxyl-terminal CAAX protein processing
 RT and amino-terminal a-factor cleavage.";
 RL Proc. Natl. Acad. Sci. U.S.A. 95:11175-11180(1998).
 CC -1- FUNCTION: PROTEOLYTICALLY REMOVES THE C-TERMINAL THREE RESIDUES OF
 CC FARNESYLATED A-FACTOR MATING PHEROMONE. ALSO ACTS TO CLEAVE THE N-
 CC TERMINAL EXTENSION OF THE PHEROMONE. DOES NOT ACT ON RAS.
 CC -1- COFACTOR: BINDS ONE ZINC ION (PROBABLE).
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. ENDOPLASMIC
 CC RETICULUM.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M48 (ZINC
 CC METALLOPROTEASE).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: 249617; CA89647.1;
 DR EMBL: 077137; AAB38271.1;
 DR MEROPS: M48.001;
 DR SGD: S0003878; STE24.
 DR InterPro: IPR001915; Peptidase_M48.
 DR InterPro: IPR000130; Zn_MTPeptide.
 DR Pfam: PF01435; Peptidase_M48; 1.
 DR PROSITE: PS00142; ZINC_PROTEASE; 1.
 KW Hydrolase; Metalloprotease; Zinc; Transmembrane;
 KW Endoplasmic reticulum; Pheromone response.
 FT TRANSMEM 13 33 POTENTIAL.
 FT TRANSMEM 90 110 POTENTIAL.
 FT TRANSMEM 122 142 POTENTIAL.
 FT TRANSMEM 168 188 POTENTIAL.
 FT TRANSMEM 198 218 POTENTIAL.
 FT TRANSMEM 307 327 POTENTIAL.
 FT TRANSMEM 358 378 POTENTIAL.
 FT METAL 297 297 ZINC (CATALYTIC) (BY SIMILARITY).
 FT ACT_SITE 298 298 BY SIMILARITY.
 FT METAL 301 301 ZINC (CATALYTIC) (BY SIMILARITY).
 FT METAL 301 301 ZINC (CATALYTIC) (BY SIMILARITY).
 FT ACT_SITE 394 394 PROTON DONOR (BY SIMILARITY).
 FT SEQUENCE 453 AA; 52324 MW; 33IC9AE2D7C99DA CRC64;
 SQ

Query Match 99.7%; Score 2336; DB 1; Length 453;
 Best Local Similarity 99.8%; Pred. No. 3.7e-160;
 Matches 452; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MEDLKTLDHPNIPMKLIISGSIAQFSSESYLYROYOKLSEKLPVLEDEIDDETH 60
 DB 1 MEDLKTLDHPNIPMKLIISGSIAQFSSESYLYROYOKLSEKLPVLEDEIDDETH 60
 QY 61 KSNINISRAKAKSIFGDVYNLAOKLVFIKYDLPKIMHNAVSLNVLVFRHMSTVAQ 120
 DB 61 KSNINISRAKAKSIFGDVYNLAOKLVFIKYDLPKIMHNAVSLNVLVFRHMSTVAQ 120
 QY 121 SICEGLSSLTVDLPISYSYSHFLEKFGFNKLTVMQMTDMKSLTLVAIGGPTL 180
 DB 121 SICEGLSSLTVDLPISYSYSHFLEKFGFNKLTVMQMTDMKSLTLVAIGGPTL 180

QY 181 YLEKIFDEKPEDFLMTYINVFVQIILMTIIPVIMPMFKPTPLEDEGLKSTESLA 240
 DB 181 YLEKIFDEKPEDFLMTYINVFVQIILMTIIPVIMPMFKPTPLEDEGLKSTESLA 240
 QY 241 DRVGFPLKIFVIDSKRSSHNAFYTGLPFSKRIVLEDTLVNSNSTDEITAVLAHEIG 300
 DB 241 DRVGFPLKIFVIDSKRSSHNAFYTGLPFSKRIVLEDTLVNSNSTDEITAVLAHEIG 300
 QY 301 HMOKNHIYVMVIFSQLHTFLFSLFTSIYRNTSFYNTGFELEKSTGSEVDVITKEPPI 360
 DB 301 HMOKNHIYVMVIFSQLHTFLFSLFTSIYRNTSFYNTGFELEKSTGSEVDVITKEPPI 360
 QY 361 IIGFNLFDLTPLECANQFVMSLSRTHRYADAVAKKLGKYNLCRALDLOIKNST 420
 DB 361 IIGFNLFDLTPLECANQFVMSLSRTHRYADAVAKKLGKYNLCRALDLOIKNST 420
 QY 421 MNVDPLSSYHSHPLEALERSALDYSEKKKN 453
 DB 421 MNVDPLSSYHSHPLEALERSALDYSEKKKN 453
 DB 421 MNVDPLSSYHSHPLEALERSALDYSEKKKN 453

RESULT 2
 ST24_SCHPO STANDARD. PRT: 474 AA.
 ID ST24_SCHPO
 AC Q10071.
 DT 01-FEB-1996 (Rel. 33, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE PROBABLE CAAX PRENYL PROTEASE 1 (EC 3.4.24.-) (PRENYL PROTEIN-
 DE SPECIFIC ENDOPROTEASE 1) (PPSEP 1).
 GN SPAC3H1.05.
 OS Schizosaccharomyces pombe (fission yeast).
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;
 OC Schizosaccharomyces.
 OX NCBI_TaxID=4896;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=972;
 RA Gentile S., Church C.M., Barrell B.G., Rajandream M.A., Walsh S.V.;
 RL Submitted (DEC-1995) to the EMBL/Genbank/DBJ databases.
 CC -1- FUNCTION: PROTEOLYTICALLY REMOVES THE C-TERMINAL THREE RESIDUES OF
 CC FARNESYLATED PROTEINS (BY SIMILARITY).
 CC -1- COFACTOR: BINDS ONE ZINC ION (POTENTIAL).
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. ENDOPLASMIC
 CC RETICULUM (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M48 (ZINC
 CC METALLOPROTEASE).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: 268144; CA92258.1;
 DR MEROPS: M48.001;
 DR InterPro: IPR001915; Peptidase_M48.
 DR InterPro: IPR000130; Zn_MTPeptide.
 DR Pfam: PF01435; Peptidase_M48; 1.
 DR PROSITE: PS00142; ZINC_PROTEASE; 1.
 KW Hypothetical protein; Hydrolase; Metalloprotease; Zinc; Transmembrane;
 KW Endoplasmic reticulum.
 FT TRANSMEM 103 123 POTENTIAL.
 FT TRANSMEM 196 216 POTENTIAL.
 FT TRANSMEM 230 250 POTENTIAL.
 FT TRANSMEM 344 364 POTENTIAL.
 FT TRANSMEM 381 401 POTENTIAL.
 FT METAL 332 332 ZINC (CATALYTIC) (BY SIMILARITY).
 FT METAL 333 333 ZINC (CATALYTIC) (BY SIMILARITY).
 FT ACT_SITE 336 336 ZINC (CATALYTIC) (BY SIMILARITY).
 FT METAL 336 336

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 12, 2002, 09:26:44 ; Search time 17.02 Seconds

(without alignments)
975.862 Million cell updates/sec

Title: US-09-165-460a-2

Perfect score: 2342
Sequence: 1 MPDLKTLIDHPNIPMKLIIS.....HPTLAERSTALDYVEKKKN 453

Scoring table:

BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 100059 seqs, 3664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database: SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|--------------|---------------------|
| 1 | 2336 | 99.7 | 453 | 1 ST24_YEAST | P47154 saccharomyc |
| 2 | 903 | 38.6 | 474 | 1 ST24_SCHPO | Q10071 schizosacch |
| 3 | 698.5 | 29.8 | 475 | 1 FACI_HUMAN | O75844 homo sapien |
| 4 | 328 | 14.0 | 426 | 1 YHPN_BACSU | P40769 bacillus su |
| 5 | 161 | 6.9 | 286 | 1 HTPX_PASMU | P57846 pasteurella |
| 6 | 150.5 | 6.4 | 310 | 1 HTPX_HELPJ | O92854 helicobacte |
| 7 | 150 | 6.4 | 335 | 1 HTPX_ARCFU | O30004 archaeoglob |
| 8 | 147.5 | 6.3 | 283 | 1 HTPX_HAEIN | P44840 haemophilus |
| 9 | 142.5 | 6.1 | 310 | 1 HTPX_HELPJ | O25582 helicobacte |
| 10 | 139 | 5.9 | 366 | 1 HTPX_AERPE | O94667 aeropyrum p |
| 11 | 138.5 | 5.9 | 258 | 1 HTPX_METTH | O26869 methanobact |
| 12 | 130.5 | 5.6 | 292 | 1 HTPX_PYRHO | O58997 pyrococcus |
| 13 | 130.5 | 5.6 | 317 | 1 HTPX_PTHAC | O94172 thermoplasma |
| 14 | 127.5 | 5.4 | 2136 | 1 YCF2_MARPO | P09975 marchantia |
| 15 | 120 | 5.1 | 292 | 1 HTPX_BUCAL | P57406 buchneria ap |
| 16 | 119.5 | 5.1 | 2294 | 1 YCF2_AARAT | P56786 arabidopsis |
| 17 | 119 | 5.1 | 297 | 1 HTPX_STRGC | O30795 streptococc |
| 18 | 116 | 5.0 | 293 | 1 HTPX_ECOLI | P23894 escherichia |
| 19 | 113 | 4.8 | 284 | 1 HTPX_METJA | O59076 methanococ |
| 20 | 113 | 4.8 | 298 | 1 HTPX_BACSU | O31657 bacillus su |
| 21 | 109 | 4.7 | 506 | 1 YCF2_PUGGR | P31921 euglena gra |
| 22 | 109 | 4.7 | 595 | 1 YCF6_METJA | O58682 methanococ |
| 23 | 107.5 | 4.6 | 1032 | 1 MT18_YEAST | P40469 saccharomyc |
| 24 | 106.5 | 4.5 | 646 | 1 NTP1_HAEPV | O37319 heliothis a |
| 25 | 106.5 | 4.5 | 1029 | 1 YFC5_YEAST | P43571 saccharomyc |
| 26 | 105.5 | 4.5 | 873 | 1 COX1_ACACA | O37370 acanthamoeb |
| 27 | 105 | 4.5 | 1427 | 1 SRB8_YEAST | P25648 saccharomyc |
| 28 | 104.5 | 4.5 | 585 | 1 YHD9_YEAST | P38732 saccharomyc |
| 29 | 104.5 | 4.5 | 1100 | 1 LOS1_YEAST | P33418 saccharomyc |
| 30 | 104 | 4.4 | 302 | 1 HTPX_AOUAE | O67798 aquifex aeo |
| 31 | 104 | 4.4 | 530 | 1 CP14_CHICK | P29760 gallus gall |
| 32 | 103.5 | 4.4 | 185 | 1 GU01_RAT | P35894 rattus norv |
| 33 | 103.5 | 4.4 | 450 | 1 VATA_SCHPO | O14265 schizosacch |

| | | | | | |
|----|-------|-----|------|--------------|--------------------|
| 34 | 103 | 4.4 | 645 | 1 VP74_NPYAC | P15963 autographa |
| 35 | 103 | 4.4 | 1101 | 1 KOES_YEAST | Q08217 saccharomyc |
| 36 | 102 | 4.4 | 577 | 1 THT1_SCHPO | O09684 schizosacch |
| 37 | 102 | 4.4 | 2493 | 1 YBA4_YEAST | P35194 saccharomyc |
| 38 | 100.5 | 4.3 | 260 | 1 YMI6_BECAM | O21266 recilmomona |
| 39 | 100.5 | 4.3 | 1040 | 1 RIK1_SCHPO | O10426 schizosacch |
| 40 | 100 | 4.3 | 1325 | 1 MRP4_HUMAN | O15439 homo sapien |
| 41 | 99.5 | 4.2 | 309 | 1 07AA_HUMAN | O76100 homo sapien |
| 42 | 99 | 4.2 | 485 | 1 YC11_KLEPN | O48457 klebsiella |
| 43 | 99 | 4.2 | 1024 | 1 RPOB_PLAFA | P21421 plasmodium |
| 44 | 98.5 | 4.2 | 313 | 1 ND2M_RHISA | O99817 rhipicephal |
| 45 | 98.5 | 4.2 | 417 | 1 O65A_DROME | P82982 drosophila |

ALIGNMENTS

| RESULT | ID | ST24_YEAST | STANDARD: | PRT: | 453 AA. |
|--------|--|--------------------------------|-----------|------|---------|
| AC | P47154: | 01-FEB-1996 (Rel. 33, Created) | | | |
| DT | 01-FEB-1996 (Rel. 33, Last sequence update) | | | | |
| DT | 20-AUG-2001 (Rel. 40, Last annotation update) | | | | |
| DE | CAAX PRENYL PROTEIN 1 (EC 3.4.24.-) (PRENYL PROTEIN-SPECIFIC | | | | |
| DE | ENDOPROTEINASE 1) (PSEEP 1) (A-FACTOR CONVERTING ENZYME). | | | | |
| GN | ST24 OR AF1 OR YJRL17M OR J2032. | | | | |
| OS | Saccharomyces cerevisiae (Baker's yeast). | | | | |
| OC | Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; | | | | |
| OC | Saccharomycetales; Saccharomycetaceae; Saccharomyces. | | | | |
| OX | NCBI_TaxID=4932; | | | | |
| RN | [1] | | | | |
| RP | SEQUENCE FROM N. A. | | | | |
| RA | Rose M., Koelter P., Ertian K. D. | | | | |
| RL | Submitted (SEP-1995) to the EMBL/GenBank/DBJ databases. | | | | |
| RN | [2] | | | | |
| RP | SEQUENCE FROM N. A., AND FUNCTION. | | | | |
| RX | MEDLINE=97167681; PubMed=9015299; | | | | |
| RA | Fujimura-Kamada K., Nouvet F. J., Michaelis S.; | | | | |
| RT | "A novel membrane-associated metalloprotease, Ste24p, is required for | | | | |
| RT | the first step of NH2-terminal processing of the yeast a-factor | | | | |
| RT | precursor."; | | | | |
| RN | [3] | | | | |
| RP | FUNCTION. | | | | |
| RX | MEDLINE=97218305; PubMed=9065405; | | | | |
| RA | Boyartchuk V. L., Ashby M. N., Rine J.; | | | | |
| RT | "Modulation of Ras and a-factor function by carboxyl-terminal | | | | |
| RT | proteolysis."; | | | | |
| RL | Science 275:1796-1800(1997). | | | | |
| RN | [4] | | | | |
| RP | FUNCTION. | | | | |
| RX | MEDLINE=9836561; PubMed=9700155; | | | | |
| RA | Tam A., Nouvet F. J., Fujimura-Kamada K., Slunt H., Sisodia S. S., | | | | |
| RT | Michaelis S.; | | | | |
| RT | "Dual roles for Ste24p in yeast a-factor maturation: NH2-terminal | | | | |
| RT | proteolysis and COOH-terminal CAAX processing."; | | | | |
| RN | [5] | | | | |
| RP | FUNCTION. | | | | |
| RX | MEDLINE=98393572; PubMed=9725832; | | | | |
| RA | Boyartchuk V. L., Rine J.; | | | | |
| RT | "Roles of prenyl protein proteases in maturation of Saccharomyces | | | | |
| RT | cerevisiae a-factor."; | | | | |
| RL | Genetics 150:95-101(1998). | | | | |
| RN | [6] | | | | |
| RP | FUNCTION. | | | | |
| RX | MEDLINE=20158928; PubMed=10692417; | | | | |
| RA | Schmidt W. K., Tam A., Michaelis S.; | | | | |
| RT | "Reconstitution of the Ste24p-dependent N-terminal proteolytic step in | | | | |
| RT | yeast a-factor biogenesis."; | | | | |
| RL | J. Biol. Chem. 275:6227-6233(2000). | | | | |
| RN | [7] | | | | |

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 12, 2002, 13:20:11 ; Search time 4622.84 Seconds

(without alignments) 10520.305 Million cell updates/sec

Title: US-09-165-460A-3
2948

Perfect score: 1 tgaactgttgatgacaagaag.....ggggagagataagaatcaca 2948

Sequence: 1 tgaactgttgatgacaagaag.....ggggagagataagaatcaca 2948

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1472140 segs, 8248589755 residues 2944280

Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: gb.ba:*
2: gb.htg:*
3: gb.in:*
4: gb.om:*
5: gb.ov:*
6: gb.pat:*
7: gb.ph:*
8: gb.pl:*
9: gb.pr:*
10: gb.ro:*
11: gb.sts:*
12: gb.sy:*
13: gb.un:*
14: gb.vl:*
15: em.ba:*
16: em.fun:*
17: em.hum:*
18: em.in:*
19: em.om:*
20: em.or:*
21: em.ov:*
22: em.pat:*
23: em.ph:*
24: em.pl:*
25: em.ro:*
26: em.sts:*
27: em.sy:*
28: em.un:*
29: em.vl:*
30: em.htgo.hum:*
31: em.htgo.inv:*
32: em.htgo.rod:*
33: em.htg.hum:*
34: em.htg.inv:*
35: em.htg.rod:*
36: em.htg.other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB | ID | Description |
|------------|--------|-------------|--------|----|-------------|-----------------------|
| C 1 | 2948 | 100.0 | 29987 | 8 | SC8156 | 249260 S. cerevisiae |
| 2 | 1560.4 | 52.9 | 4156 | 8 | D50276 | D50276 Saccharomyces |
| 3 | 1361 | 46.2 | 4130 | 8 | YSCNCR1G | I42821 Saccharomyces |
| 4 | 1163 | 39.5 | 4225 | 8 | SCU56882 | U565682 Saccharomyces |
| 5 | 1069.4 | 36.3 | 3250 | 8 | SCU32580 | U32580 Saccharomyces |
| 6 | 931 | 31.6 | 3525 | 8 | SCU63849 | U63849 Saccharomyces |
| 7 | 786.8 | 26.7 | 3240 | 8 | YSCBUL1 | D50843 Saccharomyces |
| 8 | 703.8 | 23.9 | 3020 | 8 | SCROSL1G | X88901 S. cerevisiae |
| 9 | 225.2 | 7.6 | 270 | 8 | YSCGNC58X | M87379 Yeast Eco R |
| 10 | 143.8 | 4.9 | 27559 | 8 | SC8339 | Z49210 S. cerevisiae |
| C 11 | 71.8 | 2.3 | 2023 | 8 | SCU32938 | U32938 Saccharomyces |
| C 12 | 68 | 2.3 | 4775 | 8 | SCU32938 | U32938 Saccharomyces |
| C 13 | 65.6 | 2.2 | 780 | 11 | CNS05139 | AL040561 T7 end of |
| C 14 | 65.6 | 2.2 | 256172 | 2 | AC005139 | AC005139 Plasmid |
| 15 | 65.6 | 2.2 | 310779 | 2 | AC005140 | AC005140 Plasmid |
| 16 | 64.4 | 2.2 | 904 | 11 | CNS06149 | AL399867 T3 end of |
| 17 | 64.4 | 2.2 | 86827 | 3 | PFMAL3P5 | AL034556 Plasmid |
| 18 | 63.4 | 2.2 | 1920 | 3 | PFMAL3P7 | X62393 P. falciparum |
| 19 | 63.4 | 2.2 | 110000 | 2 | PFMAL13P2_0 | AL049185 Plasmid |
| 20 | 62 | 2.1 | 253305 | 3 | PFMAL3P7 | AL034559 Plasmid |
| C 21 | 61.6 | 2.1 | 1141 | 6 | AX083744 | AL034575 Plasmid |
| 22 | 60.8 | 2.1 | 224112 | 3 | PFMAL4P2 | AL034575 Plasmid |
| 23 | 60.6 | 2.1 | 202872 | 2 | AC016160 | AC016160 Homo sapi |
| C 24 | 60 | 2.0 | 5632 | 3 | AF310892 | AF310892 Dictyostea |
| C 25 | 59.4 | 2.0 | 3214 | 3 | AF019980 | AF019980 Dictyostea |
| 26 | 59.4 | 2.0 | 318221 | 2 | PFMAL13P3 | AL049184 Plasmid |
| 27 | 58.8 | 2.0 | 321003 | 2 | PFMAL4P3 | AL035476 Plasmid |
| 28 | 58.4 | 2.0 | 78439 | 17 | HS085157 | U85157 Homo sapien |
| C 29 | 58.4 | 2.0 | 169546 | 2 | AC004157 | AC004157 Plasmid |
| 30 | 58.4 | 2.0 | 245802 | 2 | AC006279 | AC006279 Plasmid |
| 31 | 58.4 | 2.0 | 251124 | 2 | HUAE00660 | AE00660 Homo sapi |
| C 32 | 58.2 | 2.0 | 153098 | 3 | PFMAL3P2 | AL034558 Plasmid |
| 33 | 58 | 2.0 | 169794 | 2 | AC004688 | AC004688 Plasmid |
| 34 | 58 | 2.0 | 178137 | 9 | AC011302 | AC011302 Homo sapi |
| C 35 | 57.8 | 2.0 | 13684 | 3 | AE001403 | AE001403 Plasmid |
| 36 | 57.8 | 2.0 | 153477 | 2 | AC006278 | AC006278 Plasmid |
| 37 | 57.6 | 2.0 | 53932 | 2 | AC023371 | AC023371 Homo sapi |
| C 38 | 57 | 1.9 | 12029 | 3 | AE001431 | AE001431 Plasmid |
| C 39 | 57 | 1.9 | 133443 | 2 | AC006280 | AC006280 Plasmid |
| C 40 | 57 | 1.9 | 178273 | 2 | AC005308 | AC005308 Plasmid |
| C 41 | 56.8 | 1.9 | 196149 | 2 | AC004709 | AC004709 Plasmid |
| C 42 | 56.8 | 1.9 | 67970 | 3 | PFMAL1P3 | AL031746 Plasmid |
| C 43 | 56.8 | 1.9 | 149627 | 9 | AC087428 | AC087428 Homo sapi |
| C 44 | 56.6 | 1.9 | 92633 | 2 | PFMAL4P1_3 | Continuation (4 of |
| C 45 | 56.6 | 1.9 | 153477 | 2 | AC006278 | AC006278 Plasmid |

ALIGNMENTS

RESULT 1
SC8156/c * 29987 bp DNA PLN 11-AUG-1997
LOCUS S. cerevisiae chromosome XIII cosmid 8156.
DEFINITION 249260 271257
ACCESSION 249260.1 GI:809081
VERSION
KEYWORDS BUL1, cytochrome b5, DAG1, initiation factor 1A, inorganic
pyrophosphatase, IPP2, nitrate reductase, orotate
phosphoribosyltransferase, tau element, TIF1, RPS3, transfer
RNA-Gln, trehalose-phosphate synthase, U6 snRNP, URA10,
baker's yeast.
SOURCE Saccharomyces cerevisiae
ORGANISM
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; Saccharomycetaceae; Saccharomyces.
REFERENCE 1 (bases 1 to 29987)
AUTHORS Lye, G. and Churcher, C. M.
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 29987)
AUTHORS Barrett, B. and Rajandream, M. A.
TITLE Direct Submission

JOURNAL

COMMENT

Submitted (12-May-1995) *Saccharomyces cerevisiae* chromosome XIII sequencing project, Sanger Centre, Hinxton Hall, Hinxton, Cambridge CB10 1RQ E-mail: barrell@sanger.ac.uk

Notes: All CDS over 100 codons have been analysed. CDS that are completely overlapped and those that are overlapped by more than 50% of their length by a larger CDS have been omitted from this analysis.

Details of the omitted CDS are available on request. The more significant matches with motifs in the PROSITE database are also included but some of these may be fortuitous. The length in codons and the calculated codon adaptation index (CAI) is given for each CDS.

Cosmid 8156 is overlapped at the start by cosmid 9920, emb1 entry SC9920, accession no. Z48639 and at the end by cosmid 8021.

FEATURES

source

1. 29987
location/Qualifiers

/organism="Saccharomyces cerevisiae"

/strain="AB972"

/db_xref="taxon:4932"

/chromosome="XIII"

/clone="cosmid 8156"

/map="13R"

/complement(<1..356)

/note="YM8156.01c, unknown, len: 118, CAI: 0.14, overlaps and extends YM9920.13c"

/codon_start=1

/product="unknown"

/protein_id="CA88242.1"

/db_xref="GI:809082"

/db_xref="SWISS-PROT:O03496"

/translation="MTDVOFSSOEIELEFRVKEFLIANNPAKINNENKDAVLTQIEHD FRYLYIKDGLPMLNESTRLLIFPTFSICLLRSQILASKIDQSEFLSANKVELLT EANNITFEYVLEFRA"

1..140

/note="overlap with SC9920 Z48639 S. cerevisiae chromosome XIII cosmid 9920"

/complement(851..1312)

/gene="TIF11"

/complement(851..1312)

/gene="TIF11"

/note="YM8156.02c TIF11 gene, IF1A_YEAST, P3891, initiation factor 1a"

/codon_start=1

/product="TIF11p"

/protein_id="CA88243.1"

/db_xref="GI:809083"

/db_xref="SWISS-PROT:P38912"

/translation="MGKNTKGGKRGKRGKNDGPKRELLYKEEGEYVAQITKMLGN GRVASCDFGNKRNHAKIRGLKRYVMGGODIILVSLDFODDQCDVYHKYVLEART LKNGELPENAKINETDNFGESEDDVNFEGNADDEDEGEDELDDIDT"

complement(1678..4842)

/gene="TPS3"

/complement(1678..4842)

/gene="TPS3"

/note="YM8156.03c, TPS3 gene, len: 1054, CAI: 0.21, SW:TPS3_YEAST P38426, alpha, alpha-crenolose-phosphate synthase and PIR:S27471; PS00213 lipocalin signature; translation in TPS3_YEAST starts at amino acid 33 in this sequence"

/codon_start=1

/product="TPS3p"

/protein_id="CA89244.1"

/db_xref="GI:809084"

/db_xref="SWISS-PROT:P38426"

/translation="MTIIVASLFLPTPOEADVNSDPAKIVSSMTKVCNNNELS NNGERSSSVTSASHYITGLPQEAQINGEPDQRAVVGSPATGVNTNHEMELSSQFL EELTANATHAANSIGIPANNPVSGSTQRPSEVEFFSPASRVCSPQSEASASISA SRSRHHNDLSSLMKNPNLSEDSHPRVRSSSSAVITPSSKVPDPAVDVAKY REEFOQASLPMSKRVSGSTAGDSIASSSSLRYSOOFODNFIEDTSEDIDSDLE TDAIKYVVPKFGYSNNAKLPAASIMRBSYELFKHLPTIYDSDGNSIKKAVYVIA AKYVKEPVSVMGTGIPITDELPHVEYCHKIKLSDPSSFPVYVTDITFKATINIA

CDS

/complement(1678..4842)

/gene="TPS3"

/note="YM8156.03c, TPS3 gene, len: 1054, CAI: 0.21, SW:TPS3_YEAST P38426, alpha, alpha-crenolose-phosphate synthase and PIR:S27471; PS00213 lipocalin signature; translation in TPS3_YEAST starts at amino acid 33 in this sequence"

/codon_start=1

/product="TPS3p"

/protein_id="CA89244.1"

/db_xref="GI:809084"

/db_xref="SWISS-PROT:P38426"

/translation="MTIIVASLFLPTPOEADVNSDPAKIVSSMTKVCNNNELS NNGERSSSVTSASHYITGLPQEAQINGEPDQRAVVGSPATGVNTNHEMELSSQFL EELTANATHAANSIGIPANNPVSGSTQRPSEVEFFSPASRVCSPQSEASASISA SRSRHHNDLSSLMKNPNLSEDSHPRVRSSSSAVITPSSKVPDPAVDVAKY REEFOQASLPMSKRVSGSTAGDSIASSSSLRYSOOFODNFIEDTSEDIDSDLE TDAIKYVVPKFGYSNNAKLPAASIMRBSYELFKHLPTIYDSDGNSIKKAVYVIA AKYVKEPVSVMGTGIPITDELPHVEYCHKIKLSDPSSFPVYVTDITFKATINIA

gene

/complement(1678..4842)

/gene="TPS3"

/note="YM8156.03c, TPS3 gene, len: 1054, CAI: 0.21, SW:TPS3_YEAST P38426, alpha, alpha-crenolose-phosphate synthase and PIR:S27471; PS00213 lipocalin signature; translation in TPS3_YEAST starts at amino acid 33 in this sequence"

/codon_start=1

/product="TPS3p"

/protein_id="CA89244.1"

/db_xref="GI:809084"

/db_xref="SWISS-PROT:P38426"

/translation="MTIIVASLFLPTPOEADVNSDPAKIVSSMTKVCNNNELS NNGERSSSVTSASHYITGLPQEAQINGEPDQRAVVGSPATGVNTNHEMELSSQFL EELTANATHAANSIGIPANNPVSGSTQRPSEVEFFSPASRVCSPQSEASASISA SRSRHHNDLSSLMKNPNLSEDSHPRVRSSSSAVITPSSKVPDPAVDVAKY REEFOQASLPMSKRVSGSTAGDSIASSSSLRYSOOFODNFIEDTSEDIDSDLE TDAIKYVVPKFGYSNNAKLPAASIMRBSYELFKHLPTIYDSDGNSIKKAVYVIA AKYVKEPVSVMGTGIPITDELPHVEYCHKIKLSDPSSFPVYVTDITFKATINIA

CDS

/complement(1678..4842)

/gene="TPS3"

/note="YM8156.03c, TPS3 gene, len: 1054, CAI: 0.21, SW:TPS3_YEAST P38426, alpha, alpha-crenolose-phosphate synthase and PIR:S27471; PS00213 lipocalin signature; translation in TPS3_YEAST starts at amino acid 33 in this sequence"

/codon_start=1

/product="TPS3p"

/protein_id="CA89244.1"

/db_xref="GI:809084"

/db_xref="SWISS-PROT:P38426"

/translation="MTIIVASLFLPTPOEADVNSDPAKIVSSMTKVCNNNELS NNGERSSSVTSASHYITGLPQEAQINGEPDQRAVVGSPATGVNTNHEMELSSQFL EELTANATHAANSIGIPANNPVSGSTQRPSEVEFFSPASRVCSPQSEASASISA SRSRHHNDLSSLMKNPNLSEDSHPRVRSSSSAVITPSSKVPDPAVDVAKY REEFOQASLPMSKRVSGSTAGDSIASSSSLRYSOOFODNFIEDTSEDIDSDLE TDAIKYVVPKFGYSNNAKLPAASIMRBSYELFKHLPTIYDSDGNSIKKAVYVIA AKYVKEPVSVMGTGIPITDELPHVEYCHKIKLSDPSSFPVYVTDITFKATINIA

misc-feature

/complement(2260..2301)

/gene="TPS3"

/note="PS00213 lipocalin signature"

5199..6140

/note="YM8156.04, unknown, len: 313, CAI: 0.11, previously noted ORF, SW:TPS3_YEAST P38430"

/codon_start=1

/product="unknown"

/protein_id="CA89245.1"

/db_xref="GI:809085"

/db_xref="SWISS-PROT:P38430"

/translation="MNLVDAHCHVITDPDNTFGCGDGGSGGCTLCVNSNPTDNNLT KILAGRTSKNDICVGRVHPVYSHLFVYGRDRKVSHTQVLEKNEQDSDLVQVL PEPDLLEIKRENDLVISYIGIDKLFRLPANGVMYNEKALITVTKVLSHOE TVFRFRRLARHTKPSISIDVKGHGLDNCNELHYSYKILGHSYTGSKETILG OMLKRPDPPIFVLSKRNKINPKDEBDALVRSIPSCILLETIDVPIIDNPDSYOKAL TEOLOYLNAQIARAWDETLDASQALAYENFQEFIK"

6392..6997

/note="YM8156.05, unknown, len: 301, CAI: 0.12, SW:TPS3_YEAST P38429 previously noted ORF"

/codon_start=1

/product="unknown"

/protein_id="CA89246.1"

/db_xref="GI:809086"

/db_xref="SWISS-PROT:P38429"

/translation="MARPVNTAETESGRPTGCGVYASNNNGSNNNGSNNNNNN NNNNNNSNNNNNGPSSGRTNGKORLAQOQYIKRLIETHITDNDPLAKSPM DFEVYDAPLRYVDHFDLPDLTLOGYILGSLAKYTSYKRNNGQGHDKRLHKR DLAVVRRHDEHSIKETDCIPDTYKRNOKKFFKFEFG"

7278..7889

/note="YM8156.06, unknown, len: 301, CAI: 0.13, SW:TPS3_YEAST P38428, extends previously noted ORF"

/codon_start=1

/product="unknown"

/protein_id="CA89247.1"

/db_xref="GI:809087"

/db_xref="SWISS-PROT:P38428"

/translation="MDSRLITLILVGVLEFLKKFQSNQHPSSAORLSATGVNAHR PGSTONALPRTGVNGHPYTLQVETVQVLAPELHPEQRLYSTENGVSEVYERY LKGEFSEPPGFEPESRAPMGMAVDDNNAAGGEPENDRKNMTCALENITLDFBYDLN EDMNSLSEFKDLIDIERKRLVWQARKNLETLOSDKQSLT"

complement(8013..9398)

/note="YM8156.07c, unknown, len: 461, CAI: 0.11"

/codon_start=1

/product="unknown"

/protein_id="CA89248.1"

/db_xref="GI:809088"

/db_xref="SWISS-PROT:O03508"

/translation="MEERFERRRKGGSSRRGNHRYLRKANSACELAVPEVPEAI TKLHITPDNTLEATPGTEEDLYRELINLTSSITPLNGGRRNEGLOLAIEISIKTK PRYSNMADIVKLAESYSADOTISIVTGRESEFHLIEHALQATASHMGCSENEF RFNAVUCLAKRAISEYTSKRYKELMADVEYVPSLSELSIYDRIOIDAFKPSFSLD LPRKWSAIPVAPPTKSLPPEOELMADVEMVDRKNNLSQSKFQDLBRTPOIGTIL CTAARHSLIEIVKILYLRKGRRTFRPPLVHPLIYIPCAEPGKIDIPALEIAVANSND TRTDSERKVOHISQITVLEBPGACIVHPQVTLDAVYIASAHNRKRPLEYTRKAPPEP NRTFTLPEIVIGHYTKRRIEDLEVYTRLLNCKKDIEMVPLDNTPIPAEFGF AKLAIAICNSA"

9991..12852

/note="YM8156.08, unknown, len: 953, CAI: 0.20, probable integral membrane glycoprotein"

/codon_start=1

/product="unknown"

/protein_id="CA89249.1"

/db_xref="GI:809089"

/db_xref="SWISS-PROT:O03509"

/translation="MEERFERRRKGGSSRRGNHRYLRKANSACELAVPEVPEAI TKLHITPDNTLEATPGTEEDLYRELINLTSSITPLNGGRRNEGLOLAIEISIKTK PRYSNMADIVKLAESYSADOTISIVTGRESEFHLIEHALQATASHMGCSENEF RFNAVUCLAKRAISEYTSKRYKELMADVEYVPSLSELSIYDRIOIDAFKPSFSLD LPRKWSAIPVAPPTKSLPPEOELMADVEMVDRKNNLSQSKFQDLBRTPOIGTIL CTAARHSLIEIVKILYLRKGRRTFRPPLVHPLIYIPCAEPGKIDIPALEIAVANSND TRTDSERKVOHISQITVLEBPGACIVHPQVTLDAVYIASAHNRKRPLEYTRKAPPEP NRTFTLPEIVIGHYTKRRIEDLEVYTRLLNCKKDIEMVPLDNTPIPAEFGF AKLAIAICNSA"

9991..12852

/note="YM8156.08, unknown, len: 953, CAI: 0.20, probable integral membrane glycoprotein"

/codon_start=1

/product="unknown"

/protein_id="CA89249.1"

/db_xref="GI:809089"

/db_xref="SWISS-PROT:O03509"

/translation="MEERFERRRKGGSSRRGNHRYLRKANSACELAVPEVPEAI TKLHITPDNTLEATPGTEEDLYRELINLTSSITPLNGGRRNEGLOLAIEISIKTK PRYSNMADIVKLAESYSADOTISIVTGRESEFHLIEHALQATASHMGCSENEF RFNAVUCLAKRAISEYTSKRYKELMADVEYVPSLSELSIYDRIOIDAFKPSFSLD LPRKWSAIPVAPPTKSLPPEOELMADVEMVDRKNNLSQSKFQDLBRTPOIGTIL CTAARHSLIEIVKILYLRKGRRTFRPPLVHPLIYIPCAEPGKIDIPALEIAVANSND TRTDSERKVOHISQITVLEBPGACIVHPQVTLDAVYIASAHNRKRPLEYTRKAPPEP NRTFTLPEIVIGHYTKRRIEDLEVYTRLLNCKKDIEMVPLDNTPIPAEFGF AKLAIAICNSA"

9991..12852

/note="YM8156.08, unknown, len: 953, CAI: 0.20, probable integral membrane glycoprotein"

/codon_start=1

/product="unknown"

/protein_id="CA89249.1"

/db_xref="GI:809089"

/db_xref="SWISS-PROT:O03509"

/translation="MEERFERRRKGGSSRRGNHRYLRKANSACELAVPEVPEAI TKLHITPDNTLEATPGTEEDLYRELINLTSSITPLNGGRRNEGLOLAIEISIKTK PRYSNMADIVKLAESYSADOTISIVTGRESEFHLIEHALQATASHMGCSENEF RFNAVUCLAKRAISEYTSKRYKELMADVEYVPSLSELSIYDRIOIDAFKPSFSLD LPRKWSAIPVAPPTKSLPPEOELMADVEMVDRKNNLSQSKFQDLBRTPOIGTIL CTAARHSLIEIVKILYLRKGRRTFRPPLVHPLIYIPCAEPGKIDIPALEIAVANSND TRTDSERKVOHISQITVLEBPGACIVHPQVTLDAVYIASAHNRKRPLEYTRKAPPEP NRTFTLPEIVIGHYTKRRIEDLEVYTRLLNCKKDIEMVPLDNTPIPAEFGF AKLAIAICNSA"

9991..12852

/note="YM8156.08, unknown, len: 953, CAI: 0.20, probable integral membrane glycoprotein"

/codon_start=1

/product="unknown"

/protein_id="CA89249.1"

/db_xref="GI:809089"

/db_xref="SWISS-PROT:O03509"

/translation="MEERFERRRKGGSSRRGNHRYLRKANSACELAVPEVPEAI TKLHITPDNTLEATPGTEEDLYRELINLTSSITPLNGGRRNEGLOLAIEISIKTK PRYSNMADIVKLAESYSADOTISIVTGRESEFHLIEHALQATASHMGCSENEF RFNAVUCLAKRAISEYTSKRYKELMADVEYVPSLSELSIYDRIOIDAFKPSFSLD LPRKWSAIPVAPPTKSLPPEOELMADVEMVDRKNNLSQSKFQDLBRTPOIGTIL CTAARHSLIEIVKILYLRKGRRTFRPPLVHPLIYIPCAEPGKIDIPALEIAVANSND TRTDSERKVOHISQITVLEBPGACIVHPQVTLDAVYIASAHNRKRPLEYTRKAPPEP NRTFTLPEIVIGHYTKRRIEDLEVYTRLLNCKKDIEMVPLDNTPIPAEFGF AKLAIAICNSA"

9991..12852

/note="YM8156.08, unknown, len: 953, CAI: 0.20, probable integral membrane glycoprotein"

/codon_start=1

| | |
|------|--------------|
| gene | 13245..14177 |
| | /gene="IPP2" |
| CDS | 13245..14177 |

| | | | | |
|----------------------------|---------|--------------|-------|---------------|
| Query Match | 100.0%; | Score 2948; | DB 8; | Length 29987; |
| Best Local Similarity | 100.0%; | Pred. No. 0; | | |
| Matches 2948; Conservative | 0; | Mismatches | 0; | Indels 0; |

| | | | |
|----|-------|---|-------|
| OY | 481 | caacgagacacataatggaacttgaaatttaataagaaacataaagaaacttaattcc | 540 |
| Db | 27305 | CAAAAGACACATTAATATGAACTTGAAATTTAATAAGACATTAAAGAAACTTTAATTC | 27246 |
| OY | 541 | aagtttcgaagctgcctcatgtctgaagttctatctgcgtccgaatgaatcaaatctga | 600 |
| Db | 27245 | AAGTTTCAAAAGCTGCCATGTGTGAAGTTTATGTGCGTTCGAGTAAGATTAAATTGA | 27186 |
| OY | 601 | aaacacattcggcgtaagcgaagatgtataccctatcttcgttaagcgaatgcacaaata | 660 |
| Db | 27185 | AAACCATCTTGCGCGTAGCGGAAGATGAAATCCCTATTTCTGTAGGAAAGTACAAAATA | 27126 |
| OY | 661 | aaaaaaccttgaaaaaaattcgtctattttttctatatagataagatataatgtatgctt | 720 |
| Db | 27125 | AAAAAACATTGAAAAAAATTCGTTACTTTTCTTAATGATAATAATATATGATAGGTTT | 27066 |
| OY | 721 | gcttaataatgaagtatattacgcgtcccttgatctccctataatataaacttcctt | 780 |
| Db | 27065 | GCTTATATATGAGGTATTATTCGCGTCTTGTATTCCTATTTAATAAAATCTTT | 27006 |
| OY | 781 | taaatgatcttcctcgtgctcttttgttgcctctgataltttttttttttggacaactg | 840 |
| Db | 27005 | TAAATGCAATTTTCGCGCTCTTTGTGCTCGTGAATTTTTTTTTTTGGACACATG | 26946 |
| OY | 841 | gatggaaaaccttgatgataatttatacctattttaagtactcaaaatcagagattc | 900 |
| Db | 26945 | GATGGAACACCTTGATGATATTTTATACCTTATTTTAAGTACTCAAAATATCGAGATT | 26886 |
| OY | 901 | cagagacaaacataagaattctctctgcgaagaaataaaacgaaataaatgtact | 960 |
| Db | 26885 | CAGGACAAACATAGAAATTTCTTTGCAAGAAAAAATAACGAAATTAATGATGCTT | 26826 |
| OY | 961 | tgactactgactgtctgcataagagaaacgaacgaactgtcacatctccaaattc | 1020 |
| Db | 26825 | TGACTACTGACTGTGTGCATAGAGAGAACACGAAACGAATGCTATGTCATGTC | 26766 |
| OY | 1021 | tctagtctcctatacatctccatactcctatgtgtccagcgtatacgaaacttcaacc | 1080 |
| Db | 26765 | TCTATGCTCCTATACATCTCCATTCCTATGCTGACCGGTAATGCAACTTCCAAACC | 26706 |
| OY | 1081 | agaagggcttaacgaagatactcccgaaagataaactctgcatacgaaaaacttaacat | 1140 |
| Db | 26705 | AGAAAGGCTTAACGAGATAATCCGCAAGATTAAATCTCGCAAGCAAAAACTTACAT | 26646 |
| OY | 1141 | tatgctaattcccaaccctttttgtgaccttttttcaaatcccaatccatccagtaacc | 1200 |
| Db | 26645 | TATGCTAATTTCCAAACCTTTTGTGTCCTTTTAAATCTCAATATCTAGTACACAC | 26586 |
| OY | 1201 | ttcacataaagttccaagaacgaatttttagctttagatatacccaagttatacgc | 1260 |
| Db | 26585 | TTCACATTAATTAAGTTTCAAGACGCAATTTTATAGGCTTAGGTATTAATCCACAGGTTATACGC | 26526 |
| OY | 1261 | tgcattgccaaacctgtgcattccagccaattcgttgaagaacttaagaaatgtgttc | 1320 |
| Db | 26525 | TGCATTTGCCAAACCTTGCGCAATTCAGCAAGATGCTGAAGACCTTAACGAAATGTGTTGC | 26466 |
| OY | 1321 | gaatgtatctgacctatattgtgaaccggttttagatttgatattatataatcaatataa | 1380 |
| Db | 26465 | GATGTATATGACCTTAATTTGTGACCCGTTTATGATTTTGTATATATCATTTATTAA | 26406 |
| OY | 1381 | tccaaagagcctataactctgaagatctttacacatgaattcctbgaatctvgagttcag | 1440 |
| Db | 26405 | TCCAAAGAGCCTTAATCTAGTGAAGATTTTATACCATGAATTCCTGAAATTTTGGAGTTTCAG | 26346 |
| OY | 1441 | gaattttatcttcaccaaataaactgaggaatatcttaacgcgtcaaatgctttgactac | 1500 |
| Db | 26345 | GAAATTTATTTTTCGCCAAATACGAGAGAAATATTTTACACGTCGAATGCTTTACATAC | 26286 |
| OY | 1501 | gtacttaaacctataacgcattcgaactaagctatacaaacagttaatttggcaacatc | 1560 |
| Db | 26285 | GTACTTAACCTAATACCGCATTCGCAACTAGACATCAACAGTATATTTTGGCAACATC | 26226 |
| OY | 1561 | gcttttttttggacttgcgaacgaacacacatgctatagaagaattacagaagaagctcat | 1620 |

Db 26225 GCTTTTGGTGGTGGGACGACACCATCTTATGAGCAATTAACGAGAGGCTCCAT 26166
 QY 1621 gacacgtttccatctcgtcgaacacatgctccaaatttatacaaacacttttgg 1680
 Db 26165 GACAACTGTTTCATTTGCTGACACAAACATGCTTCCAAATTTTATACACAACTTTTGG 26106
 QY 1681 aggttaacacagtttctatctcgttaaacagcggtggaacatggtgtgtatattcct 1740
 Db 26105 AGGGTAACCAAGTTTGTATTCGTGAAGAACAGCGGGAACCTATGCTGATATCTT 26046
 QY 1741 gcatgccccttggaatcatcatgaggtttccgttccttcaagatggaatttcatctac 1800
 Db 26045 GCATGCCCTTTGCATATCATGAGGCTTCTGCTCTTCAATATGAAATTTTCAATTTCC 25986
 QY 1801 agtagtaacagaagaagcgtgacgcatctccaattgtcccaatctggaataagtgcta 1860
 Db 25995 AGTACTAGACAAAGAAAGCTGAGCAGCATTTCCAAATGTGCTCAATCTGAAATAAGTCTA 25926
 QY 1861 ctctgcaactgctgtcccttggaataataacccctgaaagatgtaacctctgttag 1920
 Db 25925 CTTCGCACTGCTGCTCTGATTAATATATCCCTGAAGGATACCTTAACAACCTGCTAG 25866
 QY 1921 aactcctggtatagaataaacctttagcctttttagctactgtatccgtttaaat 1980
 Db 25865 AACTCTGCTGATTAAGAAATACCTTTAGCCTTTTACGTATCTGTAATACCTTTAAAT 25806
 QY 1981 ttcccatgtactataacacttttccactactatagaaattcatcgacgacgacggtct 2040
 Db 25805 TTCCATATGATATTAACCTTTTTCCTACTATTAATGAAATCTAGACGACGACGCGCT 25746
 QY 2041 ttgttgaaggaagtgaaataacgagtttgggttttgggtgaagaatttggagagc 2100
 Db 25745 TTTGTTACCGAAGAGGAAATAATCAGATTTGGTGGTGGGAAAGATTTGGAGGAGC 25686
 QY 2101 tataaagtaaccatactctgtatctttagagtaataaacagtcgttcgttcagtgat 2160
 Db 25685 TATAAAGTACCTATATCTTTGATTAAGGACTCAATTAACAAAGTCTGCTGCTGAT 25626
 QY 2161 tgaagctgtcagatcctaaagagtagaagaagtgagcactcaataggtttcgaagctttc 2220
 Db 25625 TGAAGTGTGACATCTAAGAGTAGAAGAGAGAGAGGCGCATTAAGGTTTTCGAGCTTTTC 25566
 QY 2221 tttctttaaagtttcttactgtgtcccttaagaatttaagtcctagtttgggtgtc 2280
 Db 25565 TTTTAAAGTTTATTTTATTTGCTCTCTAATAATTTAAGGCTTGAATTTGGTGGTGT 25506
 QY 2281 ttgtgtgtatcatatttcaattcaagaagaatttagtgcctttataatggtccaat 2340
 Db 25505 TTTGTGGGTACATATTTTCAATTCAAAGAGAAATTTAGCTGTCTTTTATTAATGCTCAAT 25446
 QY 2341 agagtaacagagagcgtcgtctactacatacgaagcagataagcgatcgtctagtcgaag 2400
 Db 25445 AAGGATTAACGAGAGCATGCGCTGCTACTACATCAAGGATTAAGGAGCGGTATGCAAAAG 25386
 QY 2401 gataaagcgaagtcgaagtttgcattgtgcacagtcctctgacaaatgaatccgagc 2460
 Db 25385 GATAAAGCGAACTGAAGTTTGAATGCTGCAAGAGCTCTTGCAATGAATAATCCGAGC 25326
 QY 2461 gtaaaaaactaaagattgtcattgtgttgatgagatttactattatgacgaatla 2520
 Db 25325 GTAATAAACTAAAGATTGTGATGAGTCAATGATTTACTATTGATCCAGAAATTA 25266
 QY 2521 gataaaaaactcgtgtggaatcgtatggtggaagacgacatggtctgacagacatccagt 2580
 Db 25265 GATATTAATTCGCTGGGGAATCTAGTGGAGAGAGCATGAGTGTGACAGACATCCAGT 25206
 QY 2581 tctgtcgtcaatgccaagtgcacacacacacccgttaataacacagataatagcgtaccact 2640
 Db 25205 TCTGCGCATATGCAAGTGAACAACACCGTTAATAACACAGATATAGGAGTCAACT 25146
 QY 2641 ccgtcagagaagacttgatggagggttaactcaaggatagaatccctccaaataagactaa 2700

Db 25145 CCGCTAGAACTTGATGAGGAGGCTAATCAGGATAGAAATCCTCCATTAAGACTAA 25086
 QY 2701 caagtaactacttaggtatataaaagtggttcaactccatccaggaataatgct 2760
 Db 25085 CAAGTACTACTTATGATTAATAAAGAGGCTTCTCTCATCCAGGAAATTAATGCT 25026
 QY 2761 aacgtataaagaataacttatagggttcccgcaatcaaacaccttaagct 2820
 Db 25025 AACGATTAAGAAATAATTTATGAGTTCCTCCGCAATCAACACCTTAAGCT 24966
 QY 2821 gataatctccttagagctgttcaagaataccttatacaataacaaagcgaatgct 2880
 Db 24965 GATAATTTCTAGAGCTTGTAAGATCTTTTACAAATATACAACTAAGCAGCATGCT 24906
 QY 2881 gaagtaatggtggaataagaataaataagcatttgaggaatgaggaggaataa 2940
 Db 24905 GAAGTAATGATGGAATAGCAATGAATAAATGATTTGAGGATTAATGGAGATTA 24846
 QY 2941 gaacaca 2948
 Db 24845 GATACCA 24838

RESULT 2

D50276 4156 bp DNA PLN 10-FEB-1999
 LOCUS Saccharomyces cerevisiae HST1 gene, complete cds.
 DEFINITION
 ACCESSION D50276.1 GI:1777313
 VERSION HST1; high-copy suppressor of TFP sensitivity.
 KEYWORDS Saccharomyces cerevisiae (strain:RAY-3A) DNA.
 SOURCE Saccharomyces cerevisiae
 ORGANISM Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; Saccharomycetales; Saccharomycetaceae; Saccharomyces.

REFERENCE
 AUTHORS Tsuchiya, E.
 TITLE Submitted (13-Apr-1995) to the DDBJ/EMBL/GenBank databases. Eiko Tsuchiya, Hiroshima University Faculty of Engineering, Department of Fermentation Technology, Kagami-yama, Higashi-Hiroshima, Hiroshima 739, Japan (E-mail:etsuchi@ipc.hiroshima-u.ac.jp, Tel:0824-24-7868, Fax:0824-24-7869).
 JOURNAL Unpublished (1995)

REFERENCE
 AUTHORS Tsuchiya, E., Matsuzaki, G., Tsukao, A. and Miyakawa, T.
 TITLE Molecular cloning and characterization of S. cerevisiae HST1 gene
 JOURNAL Unpublished (1995)

REFERENCE
 AUTHORS Tsuchiya, E., Matsuzaki, G., Kuran, K., Fukuchi, T., Tsukao, A. and Miyakawa, T.
 TITLE The Saccharomyces cerevisiae SSD1 gene is involved in the tolerance to high concentration of Ca²⁺ with the participation of
 JOURNAL HST1/NRC1/BF1
 MEDLINE Gene 176 (1-2), 35-38 (1996)
 FEATURES 97075905

FEATURES

source
 1..4156
 /organism="Saccharomyces cerevisiae"
 /strain="RAY-3A"
 /db_xref="taxon:4932"
 946..3693
 /gene="HST1"
 946..3693
 /gene="HST1"
 /standard_name="high-copy suppressor of TFP sensitivity"
 /codon_start=1
 /protein_id="BAA08819.1"
 /db_xref="GI:1816450"

CDS

/translation="MSNRNESHMLRTSSDKAIAQRDRKSPVLAASIDNEIRSY
 KNLKRLSHGNGSMDLLIDPELDIFGSGSRKSGSTSSASHPSTTVNATRIDP
 PLEMLHGRGNGSIESNKTQGNVGIKGVHSPKRNANLAKNLLVNPANQHPN
 VPDNLEIYQDPTLQISDNGEDNGDSNENNDIEDNGEDSOSYERKENTINL
 NRGLSRHGNASLIRPSTIRRSYTERDNDNDNDKDSASETVNRYEERSTIKERY
 SRDITELTKISNSAGLTNDNDAITLARIASGISTDKDQRPFGHDEGDIIGST

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 12, 2002, 10:38:12 ; Search time 4622.84 Seconds

(without alignments)
6512.739 Million cell updates/sec

Title: US-09-165-460A-1
1825

Perfect score: 1 accactctttttctatct.....aaaataacgattaacatt 1825

Sequence: 1 accactctttttctatct.....aaaataacgattaacatt 1825

Scoring table: IDENTITY_NTC
Gapop 10.0 ; Gapext 1.0

Searched: 1472140 segs, 824899755 residues 2944280

Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Database :

Listing first 45 summaries

GenEmbl:
1: gb_ba:*
2: gb_hgt:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_om:*
20: em_or:*
21: em_ov:*
22: em_pat:*
23: em_ph:*
24: em_pl:*
25: em_ro:*
26: em_sts:*
27: em_sy:*
28: em_un:*
29: em_vl:*
30: em_hgt_hum:*
31: em_hgtgo_iny:*
32: em_hgtgo_rtd:*
33: em_hgt_hum:*
34: em_hgt_iny:*
35: em_hgt_rtd:*
36: em_hgtg_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB | ID | Description |
|------------|--------|-------------|--------|----|-----------|----------------------|
| 1 | 1823.4 | 99.9 | 1825 | 8 | SCYJR117W | 249617 S. cerevisiae |
| 2 | 1553.4 | 85.1 | 1706 | 8 | SCU77137 | U77137 Saccharomyc |
| 3 | 342 | 18.7 | 1555 | 8 | SCYJR116W | 249616 S. cerevisiae |
| 4 | 195 | 10.7 | 99360 | 8 | NCHL1N2 | AL513444 Neurospor |
| 5 | 187 | 10.2 | 36920 | 8 | SPAC3H1 | 268144 S. pombe ch |
| 6 | 137.2 | 7.5 | 1275 | 8 | AF353722 | AF353722 Arabidops |
| 7 | 136 | 7.5 | 686 | 6 | A86270 | A86270 Sequence 92 |
| 8 | 136 | 7.5 | 686 | 6 | ARI55763 | ARI55763 Sequence |
| 9 | 136 | 7.5 | 686 | 6 | E66288 | E66288 genome DNA |
| 10 | 121 | 6.6 | 935 | 8 | SCYJR118C | 249618 S. cerevisiae |
| 11 | 117.4 | 6.4 | 1859 | 9 | AF064867 | AF064867 Homo sapi |
| 12 | 117.4 | 6.4 | 2963 | 9 | AB016068 | AB016068 Homo sapi |
| 13 | 117.4 | 6.4 | 2966 | 9 | HSY13834 | Y13834 Homo sapien |
| 14 | 117.4 | 6.4 | 2968 | 6 | E32056 | E32056 Human AFCL |
| 15 | 114.2 | 6.3 | 1340 | 3 | AF358443 | AF358443 Physarum |
| 16 | 98.4 | 5.4 | 78592 | 3 | AC004335 | AC004335 Drosophila |
| 17 | 98.4 | 5.4 | 102413 | 2 | AC020024 | AC020024 Drosophila |
| 18 | 92.8 | 5.1 | 41327 | 3 | CBRG47J19 | CBRG47J19 Caenorhab |
| 19 | 91.6 | 5.0 | 40767 | 3 | CRC04F12 | 281461 Caenorhabd |
| 20 | 88 | 4.8 | 317511 | 1 | CJ11168X3 | AL139076 Campyloba |
| 21 | 86.6 | 4.7 | 582 | 8 | CNS018P | AL110644 Botrytis |
| 22 | 81.6 | 4.5 | 660 | 8 | CNS018A3 | AL112927 Botrytis |
| 23 | 72.2 | 4.0 | 2161 | 9 | AK027874 | AK027874 Homo sapi |
| 24 | 66.8 | 3.7 | 1072 | 6 | AX055334 | AX055334 Sequence |
| 25 | 66.6 | 3.6 | 124808 | 2 | AP000652 | AP000652 Homo sapi |
| 26 | 65.6 | 3.6 | 16488 | 1 | AE004992 | AE004992 Homo sapi |
| 27 | 65 | 3.6 | 256172 | 2 | AC005139 | AC005139 Halobacte |
| 28 | 64.4 | 3.5 | 169423 | 2 | AC069309 | AC069309 Mus muscu |
| 29 | 64.4 | 3.5 | 186306 | 2 | AC083946 | AC083946 Mus muscu |
| 30 | 63.6 | 3.5 | 1245 | 6 | AX055202 | AX055202 Sequence |
| 31 | 63.6 | 3.5 | 19715 | 1 | AE000555 | AE000555 Helicobac |
| 32 | 62.4 | 3.5 | 234112 | 3 | PFMALP2 | AL035475 Plasmodiu |
| 33 | 62.4 | 3.4 | 13383 | 3 | AE001374 | AE001374 Plasmodiu |
| 34 | 61.8 | 3.4 | 98734 | 2 | PFMALP2 | AL031745 Plasmodiu |
| 35 | 61.6 | 3.4 | 14943 | 1 | AE001528 | AE001528 Helicobac |
| 36 | 59.8 | 3.3 | 12029 | 3 | AE001392 | AE001392 Plasmodiu |
| 37 | 59.8 | 3.3 | 110000 | 2 | AC087899 | AC087899 Mus muscu |
| 38 | 59.2 | 3.2 | 114736 | 3 | PFMALP3 | 298547 Plasmodiu |
| 39 | 59.2 | 3.2 | 253505 | 3 | PFMALP3 | AL034559 Plasmodiu |
| 40 | 58.8 | 3.2 | 98899 | 2 | AL603743 | AL603743 Danio rer |
| 41 | 58.8 | 3.2 | 318221 | 2 | PFMALP3P3 | AL049184 Plasmodiu |
| 42 | 58.8 | 3.2 | 318503 | 2 | AC074166 | AC074166 Mus muscu |
| 43 | 58.4 | 3.2 | 172239 | 2 | AC069587 | AC069587 Homo sapi |
| 44 | 58.2 | 3.2 | 12029 | 3 | AE001412 | AE001412 Plasmodiu |
| 45 | 58.2 | 3.2 | 187469 | 9 | AL357312 | AL357312 Human DNA |

ALIGNMENTS

RESULT 1
SCYJR117W 1825 bp DNA PLN 11-AUG-1997
LOCUS S. cerevisiae chromosome X reading frame ORF YJR117W.
DEFINITION
ACCESSION Z49617.1 Y13136
VERSION Z49617.1 GI:1015836
KEYWORDS
SOURCE
ORGANISM

baker's yeast.
Saccharomyces cerevisiae

Eukaryotes; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; Saccharomycetaceae; Saccharomyces.

REFERENCE 1 (bases 1 to 1825)
AUTHORS Rose, M., Koetter, P. and Entian, K.D.

JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1825)
AUTHORS MIPS.

TITLE Direct Submission

Submitted (25-SEP-1995) Data collected by MIPS on behalf of the European yeast chromosome X sequencing project. MIPS at the Max-Planck-Institut fuer Biochemie, Am Klopferspitze 18a D-82152

| | | | |
|----|-----|---|-----|
| QY | 1 | accaccctttttctctctcctcaacgaagcctctcaacacacacataact | 60 |
| Db | 1 | ACCTACCTTTTTTCTCFATCTTCAACAAGAAACGCTTACACACACATACATCT | 60 |
| QY | 61 | acataacatacctacaatatatacatatgtaacttgtatatattccctattaaccaa | 120 |
| Db | 61 | ACATACATACCTATACAAATATACATATATGTAAACTTGTATATTCCTTAAACCAA | 120 |
| QY | 121 | aaggagcaattaaactttccctctttttctagtcacttaactcaaaactcaattcct | 180 |
| Db | 121 | AAGAGCAATTAATTAATCTTTCCTCTTTTCTACGTCATTTACTCAAAACTCTAATTCCT | 180 |
| QY | 181 | tcgtctctgtctgcacatttccccaagaaaaaatcgacgggaaataaaaaaaagac | 240 |
| Db | 181 | TGCTCTCTGTTCTCGCATTTTCTCCAGAAAAAATCGACGGGAATAAAAAAGAC | 240 |
| QY | 241 | aacgaacaaggaaaaagttcgcgaattatataaacaccttcataattaaacgaagaa | 300 |
| Db | 241 | AACGAACAAGAAAAAGTTGCGGATATATTAACCACTTCTATATTACCGGAAAAAGGA | 300 |
| QY | 301 | aggaaaaaaaaggagaatagaataaactcgacggccttatcatcgttgaacttaagc | 360 |
| Db | 301 | AGGAAAAAAAGAGGAATATGAATAACGACGGCCTTATTCAGTGTGATCTTAACG | 360 |
| QY | 361 | attctcgacatccataatcccgltgaaattatcatcttcgttggttcgatgccc | 420 |
| Db | 361 | ATTCTCGACATCTCTTAATATCCCGGAAATTAATCATTTCTGGGTTCTCGATTGCCAA | 420 |
| QY | 421 | ttttcttcgatacttaacttgaactgacgaagaatccgaagcctatctgaaacaagt | 480 |
| Db | 421 | TTTTCTTTCGATCTTAATTGACGACGACAGACAGAAAGCATCTGAAACAAATTG | 480 |
| QY | 481 | ccactctgctgcgaagaagaattatgatagaactttctaataatcaagaactactcc | 540 |
| Db | 481 | CCACTCTGCTGGAAGACGAATTGATATGAATTTTCTAATAATCAAGAACTTACTCC | 540 |
| QY | 541 | cggcgcaagcgcaagtctccatttcgttgcgtctataaccctagccccaaagctagt | 600 |
| Db | 541 | CGGCGCAAGGCGCAAGTCTCCATTTCGTTGCTAGCGCTATAACCTACGCCAAAACCTAGTT | 600 |
| QY | 601 | ttcatccaatacgaactctccctcaaaactctgcacatgycgcgtttccttatgaatga | 660 |

| | | | |
|----|------|--|------|
| Dh | 601 | TTATCATAAATAGACCTCTTCCCTAAATCTGGCAATAGGCCGTTCTTTATTAATGA | 660 |
| Qy | 661 | gtccctgcagtcagattcaattcaattgctccacgtctcgcaagagttatgtcttgggt | 720 |
| Dh | 661 | GTCTCCCGCATGATGATTTCAATGATGCTCCACCTGCGACAGAGATTAATGTTCTTGGGT | 720 |
| Qy | 721 | ctctttccaaattttgtctaaccttggtttgatttgcaactcttacttaagcaatttggc | 780 |
| Dh | 721 | CTCTTATCCAGTTTGTGCTACCTTGATGATTAATGGCACCTCTTACTATAGCAATTTTGTG | 780 |
| Qy | 781 | ctggaagaaaaaatttggtttccaataatctgacccgtccacactatgataccagatgac | 840 |
| Dh | 781 | CTGGAGAGAAAAATTTGGTTTCAATAAATTGACCGGTCCAACTATGAGATCACCGATATGATC | 840 |
| Qy | 841 | aagaatctgactttggcgtatgctatttgggtggcccaatcccttaccgttcccttaagac | 900 |
| Dh | 841 | AAGAGCTACTACTTTGGCGATGCTATTTGGTGGCCCAATCTTTACGTTCTTTAAGATC | 900 |
| Qy | 901 | tttgataaaattccctacgtatcttctcttggatcaataagtcctctgttcgtgtccaa | 960 |
| Dh | 901 | TTTGATATAAATTTCCCTACTCTATTTCTTTGGTGCATTAATAGCTCTTCTTGTGTGTCCA | 960 |
| Qy | 961 | atcttagccaatgacaatcatcattccagctcttcatactgcccacgttttaabaagtcactca | 1020 |
| Dh | 961 | ATCTTAGCCATGACAAATCATCTCCAGTCTTTCATGCAATGCCATGTTTAATAAGTCACTCA | 1020 |
| Qy | 1021 | ttggaggaagcgtggaactgaaaaaactcttgaagtttggccgatagagttgggttccct | 1080 |
| Dh | 1021 | TTGGAGGAGCGGTGAATGTGAAAAAATCTATTTGAAATTTGGCGGATAGAGTTGGGTTCCCT | 1080 |
| Qy | 1081 | ctgataaagattttgtcatcttgaacggcccaaaaagaatttccaattccaagaactatttc | 1140 |
| Dh | 1081 | CTGATAAAGATTTTGTCTCATTTGACGCTCAAAAAAATTTCTTCATCTAAAGCAATATTTTC | 1140 |
| Qy | 1141 | acagatttgcacatcccccacaagaagaattgttggttcgcacacttgaagaaagtaat | 1200 |
| Dh | 1141 | ACAGATTTGGCAATCACTCCCAAGAGAAATGTTTGTTCGACACTTATAGTAAGATAT | 1200 |
| Qy | 1201 | tctactgatagaattcaaggcgttlttggcccaatgaatccggtcgaacaaaataaacac | 1260 |
| Dh | 1201 | TCTACTGATGAATTAATTCGGGCTGTTTGGCCCAATGAGGCTCATGCGCAAAAAAACAC | 1260 |
| Qy | 1261 | atgttgtaaatgttactctttagtcaatttgcaacacttcccatcttctccctttcac | 1320 |
| Dh | 1261 | ATGCTTAATATGGTCAATCTTAGTCAATATGGCACACTTCTCATTTTCTCCCTTTTCAC | 1320 |
| Qy | 1321 | agcactcagaataatcatcattttacaacaaccttcggcttcttcttagaagaatgccact | 1380 |
| Dh | 1321 | AGCATCTACGAATAATCATCATTTTACAACACCTTGCGCTTTTCTTAGAGAATGCCACT | 1380 |
| Qy | 1381 | ggcagtttgtgataccgttcaactaaagaaattcccaatbatcatgtgacttaagtta | 1440 |
| Dh | 1381 | GGCAGTTTGTGATCCCGTATATCACTAAGAAATTCGCCCATATCATTTAGATTTATGTTA | 1440 |
| Qy | 1441 | tttaagactataatcctccactcgtgaatgtgcacagcaatccgtgatgagttaattccc | 1500 |
| Dh | 1441 | TTTAAGCACTAATACTCCACTCGAATGGCAATGGCAATTCGATAGTTTAATTTCC | 1500 |
| Qy | 1501 | agaacatagaataatacaagcgtatgcttagtctaaaaaattgggtctacaagcaaaactca | 1560 |
| Dh | 1501 | AGAACTATATAATACAGCTGATGCTTATGCTAAAAAATTGGCTACAAAGCAAAATCTA | 1560 |
| Qy | 1561 | tttagggcttaattgatctacaataatcaaaaaccttccacacatgaatgtatgactcttg | 1620 |
| Dh | 1561 | TGTAGGGCTTATTAATGATCTACAAATCAAAAACCTTTCCACCATGAATAGTATGATCTGTG | 1620 |
| Qy | 1621 | tattctagctataattatcccatccaactctagctgaagaatgcacgctcttagactat | 1680 |
| Dh | 1621 | TATCTAGCTATCATTAATTTCCATCCCAACTTAAGTGAAGAATGACCGCTCTAGACTAT | 1680 |
| Qy | 1681 | gttagtgaaaaagaanaaactaatctctatagatgatacatatagcatgtaccgtttaat | 1740 |
| Dh | 1681 | GTTAGTGAAGAAAGAAAACCTAATCTCTATAGATGATACATATAGCATGTACCGTTTAAT | 1740 |

| | | | |
|----|------|---|------|
| QY | 331 | aggccttatctatgcttgcattcctaagaacgcttcgcacacccctccataatcccgctggaaa | 390 |
| Db | 61 | aggcctttatctatgcttgcattcctaagaacgcttcgcacacccctccataatcccgctggaaa | 120 |
| QY | 391 | ttaatacttccgggtcttcgcagtccccaattctcttcgaaactctaacctgaactagaca | 450 |
| Db | 121 | tttaatacttccgggtcttcgcagtccccaattctcttcgaaactctaacctgaactagaca | 180 |
| QY | 451 | caagttaccagaagctatctgaaacaaagtctgcaccctctgctcgtggaagaaatgtgat | 510 |
| Db | 181 | caagttaccagaagctatctgaaacaaagtctgcaccctctgctcgtggaagaaatgtgat | 240 |
| QY | 511 | gaaactttcctaataatcaagaagaaactactccggcgccaaagcttccatttgcgt | 570 |
| Db | 241 | gaaacttttcatataaacaagaagaaactactccggcgccaaagcttccatttgcgt | 300 |
| QY | 571 | gaagctcataaacttagcccaaaagctagtttttcatcaaatcgaacttccccaataac | 630 |
| Db | 301 | gaagctcataaacttagcccaaaagctagtttttcatcaaatcgaacttccccaataac | 360 |
| QY | 631 | tggcaaatggccgtttcccttattgaaatgcagtcgcctgcagctagaattcatalgtctcc | 690 |
| Db | 361 | tggcaaatggccgtttcttattgaaatgcagtcgcctgcagctagaattcatalgtctcc | 420 |
| QY | 691 | actgcgcacagaagttatgcgtctctggctctctatccagttgtctactcgttgat | 750 |
| Db | 421 | actgcgcacagaagttatgcgtctctggctctctatccagttgtctactcgttgat | 480 |
| QY | 751 | ttgcacactccttaccatgatggccattttgcccgtggaagaaaaattggtttcaataatg | 810 |
| Db | 481 | ttgcacactccttaccatgatggccattttgcccgtggaagaaaaattggtttcaataatg | 540 |
| QY | 811 | accgtccaactatgatacacaagatgatcaagaagctcgaacttggctatgcatttgt | 870 |
| Db | 541 | accgtccaactatgatacacaagatgatcaagaagctcgaacttggctatgcatttgt | 600 |
| QY | 871 | ggcccaactccttaccctgttcccttaagaactcttgataaattccctactgattcccttg | 930 |
| Db | 601 | ggcccaactccttaccctgttcccttaagaactcttgataaattccctactgattcccttg | 660 |
| QY | 931 | taactatgctctctctgttctgttcgccaactcttagccatgacaataatcccaagcttc | 990 |
| Db | 661 | taactatgctctctctgttctgttcgccaactcttagccatgacaataatcccaagcttc | 720 |
| QY | 991 | atcatgcacatgcttctaataagttcacctccacttggagagacgctgaaactgaaactat | 1050 |
| Db | 721 | atcatgcacatgcttctaataagttcacctccacttggagagacgctgaaactgaaactat | 780 |
| QY | 1051 | gaaagtctggcagatagagttgggtctccctctagataaagattttgtcatgtacgctca | 1110 |
| Db | 781 | gaaagtctggcagatagagttgggtctccctctagataaagattttgtcatgtacgctca | 840 |
| QY | 1111 | aaaaagatcttctatccaagaagataattccaaggtttgcacttcaactccaagaagaat | 1170 |
| Db | 841 | aaaaagatcttctatccaagaagataattccaaggtttgcacttcaactccaagaagaat | 900 |
| QY | 1171 | gtttgtctgcacacttagtgaacagtaattctacatgaaatcaagcctgtttgcgc | 1230 |
| Db | 901 | gtttgtgtgcacacttagtgaacagtaattctacatgaaatcaagcctgtttgcgc | 960 |
| QY | 1231 | catgaatcgtgcacatgycacaaaaaacacacatcgttaataatgtaacatcttagtcaatg | 1290 |
| Db | 961 | catgaatcgtgcacatgycacaaaaaacacacatcgttaataatgtaacatcttagtcaatg | 1020 |
| QY | 1291 | caacacctctcatatttctccctttccaccagacatctcaagaataataatcatttaacac | 1350 |
| Db | 1021 | caacacctctctcatatttctccctttccaccagacatctcaagaataataatcatttaacac | 1080 |
| QY | 1351 | acctctggcctttctcttaagaagaagctcactgacgtttgtgtgacccgtttacaaga | 1410 |
| Db | 1081 | acctctggcctttttcttaagaagaagctcactgacgtttgtgtgacccgtttacaaga | 1140 |
| QY | 1411 | gaattcccaatcatcttgattatgtaatttaagaactttaactccaactgaaagt | 1470 |

THIS PAGE BLANK (13PT0)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 12, 2002, 10:38:12 ; Search time 4622.84 Seconds
(without alignments)
6512.739 Million cell updates/sec

Title: US-09-165-460A-1
Perfect score: 1825
Sequence: 1 acctaccctttttttctatctc.....aaataaagataaacatt 1825

Scoring table: IDENTITY_NDC
Gapop 10.0 , Gapext 1.0

Searched: 1472140 segs, 8248589755 residues
Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl.*
1: gb_ba.*
2: gb_htg.*
3: gb_in.*
4: gb_cm.*
5: gb_ov.*
6: gb_pat.*
7: gb_ph.*
8: gb_pl.*
9: gb_pr.*
10: gb_ro.*
11: gb_sts.*
12: gb_sy.*
13: gb_un.*
14: gb_vl.*
15: em_ba.*
16: em_fun.*
17: em_hum.*
18: em_in.*
19: em_com.*
20: em_ov.*
21: em_or.*
22: em_pat.*
23: em_ph.*
24: em_pl.*
25: em_ro.*
26: em_sts.*
27: em_sy.*
28: em_un.*
29: em_vl.*
30: em_htgo_hum.*
31: em_htgo_inv.*
32: em_htg_hum.*
33: em_htg_inv.*
34: em_htg_inv.*
35: em_htg_inv.*
36: em_htg_inv.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|--------|-------------|--------|--------------|---------------------|
| 1 | 1823.4 | 99.9 | 1825 | 8 SCYR117W | 249617 S.cerevisia |
| 2 | 1553.4 | 85.1 | 1706 | 8 SCYR7137 | 077137 Saccharomy |
| 3 | 342 | 18.7 | 1555 | 8 SCYR116W | 249616 S.cerevisia |
| 4 | 195 | 10.7 | 99360 | 8 NCBI112 | AL513444 Neurospor |
| 5 | 187 | 10.2 | 36920 | 8 SPAC3H1 | 268144 S. pombe ch |
| 6 | 137.2 | 7.5 | 1275 | 8 AF353722 | AF353722 Arabidops |
| 7 | 136 | 7.5 | 686 | 6 AB6270 | A86270 Sequence 92 |
| 8 | 136 | 7.5 | 686 | 6 AR155763 | AR155763 Sequence |
| 9 | 136 | 7.5 | 686 | 6 E66288 | E66288 Genome DNA |
| 10 | 121 | 6.6 | 935 | 8 SCYR118C | 249618 S.cerevisia |
| 11 | 117.4 | 6.4 | 1859 | 9 AF064867 | AF064867 Homo sapi |
| 12 | 117.4 | 6.4 | 2963 | 9 AB016068 | AB016068 Homo sapi |
| 13 | 117.4 | 6.4 | 2966 | 9 HS13834 | Y13834 Homo sapien |
| 14 | 117.4 | 6.4 | 2968 | 6 E32056 | E32056 Human AF1 |
| 15 | 114.2 | 6.3 | 1240 | 3 AF358443 | AF358443 Physarum |
| 16 | 98.4 | 5.4 | 78592 | 3 AC004335 | AC004335 Drosophila |
| 17 | 98.4 | 5.4 | 102413 | 2 AC020024 | AC020024 Drosophila |
| 18 | 92.8 | 5.1 | 41327 | 3 CBRG47J19 | AC084676 Caenorhab |
| 19 | 91.6 | 5.0 | 40767 | 3 CEC04812 | Z81461 Caenorhabd |
| 20 | 88 | 4.8 | 317511 | 1 C111168X3 | AL139076 Campyloba |
| 21 | 86.6 | 4.7 | 582 | 8 CNS018P | AL110644 Botrytis |
| 22 | 81.6 | 4.5 | 660 | 8 CNS018P3 | AL112927 Botrytis |
| 23 | 72.2 | 4.0 | 2161 | 9 AK027874 | AK027874 Homo sapi |
| 24 | 66.8 | 3.7 | 1072 | 6 AX055334 | AX055334 Sequence |
| 25 | 66.6 | 3.6 | 124808 | 2 AP000652 | AP000652 Homo sapi |
| 26 | 65.6 | 3.6 | 16488 | 1 AE004992 | AE004992 Halodactyl |
| 27 | 65 | 3.6 | 256172 | 2 AC005139 | AC005139 Plasmodiu |
| 28 | 64.4 | 3.5 | 169423 | 2 AC069309 | AC069309 Mus muscu |
| 29 | 64.4 | 3.5 | 186306 | 2 AC083946 | AC083946 Mus muscu |
| 30 | 63.6 | 3.5 | 1245 | 6 AX055202 | AX055202 Sequence |
| 31 | 63.6 | 3.5 | 19715 | 1 AE000555 | AE000555 Helicobac |
| 32 | 63.6 | 3.5 | 234112 | 3 PFMA14P2 | AL035475 Plasmodiu |
| 33 | 62.4 | 3.4 | 13383 | 3 AE001374 | AE001374 Plasmodiu |
| 34 | 61.8 | 3.4 | 98734 | 2 PFMA11P2 | AL031745 Plasmodiu |
| 35 | 61.6 | 3.4 | 14943 | 1 AE001528 | AE001528 Helicobac |
| 36 | 59.8 | 3.3 | 12029 | 3 AE001392 | AE001392 Plasmodiu |
| 37 | 59.8 | 3.3 | 110000 | 2 AC087899_0 | AC087899 Mus muscu |
| 38 | 59.2 | 3.2 | 114736 | 3 PFMA13P3 | Z98547 Plasmodium |
| 39 | 59.2 | 3.2 | 253505 | 3 PFMA13P7 | AL034559 Plasmodiu |
| 40 | 58.8 | 3.2 | 98899 | 2 AL603743 | AL603743 Danto rer |
| 41 | 58.8 | 3.2 | 318221 | 2 PFMA13P3 | AL049184 Plasmodiu |
| 42 | 58.8 | 3.2 | 318503 | 2 AC074166 | AC074166 Mus muscu |
| 43 | 58.4 | 3.2 | 172239 | 2 AC069587 | AC069587 Homo sapi |
| 44 | 58.2 | 3.2 | 12029 | 3 AE001412 | AE001412 Plasmodiu |
| 45 | 58.2 | 3.2 | 187469 | 9 AL357312 | AL357312 Human DNA |

ALIGNMENTS

RESULT 1
SCYR117W 1825 bp DNA PLN 11-NOV-1997
LOCUS S.cerevisiae chromosome X reading frame ORF YXR117W.
DEFINITION Z49617 Y13136
ACCESSION Z49617.1 GI:1015836
VERSION
KEYWORDS
SOURCE
ORGANISM

Saccharomyces cerevisiae

Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;

REFERENCE 1 (bases 1 to 1825)
AUTHORS Rose, M., Koetter, P. and Entian, K. D.

JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1825)
AUTHORS MIPS.

TITLE Direct Submission
JOURNAL Submitted (25-SEP-1995) Data collected by MIPS on behalf of the European Yeast Chromosome X sequencing project. MIPS at the Max-Planck-Institut fuer Biochemie, Am Klopfersplatz 18a D-82152

Qy 271 aaaccacttctataatcacaggaagaagaaaaaagaggaatagaaactgc 330
|||||
Db 1 AAACCACTTCTATATTAAACAGAAAGAGAAATAAAGAGAGAAATAGAAACTGC 60

| | | | |
|----|------|---|------|
| Oy | 331 | aggccttatactatgtttagatcttbaagcagatcttcggccactctctaataatccgttgaa | 390 |
| Db | 61 | aggcctttatctatggtttatgtttaaagacgatttcgacacatccaaatcccgtaa | 120 |
| Oy | 391 | ttaacatcttcgttgcttcgaatgacccaaattctcttcgaactctactaagctagacaga | 450 |
| Db | 121 | tttaatcattttcgttggttcgcgattgcgccaaattttcttgcgaattcttaactgactacaga | 180 |
| Oy | 451 | cagtaacagagaactctgaaacaaagtctgcaccctgtgcctggaagacgaattgatgat | 510 |
| Db | 181 | cagtaacagagaactatctgaaacaaagtctgcaccctgtgcctggaagacgaattgatgat | 240 |
| Oy | 571 | gactctcataactagcccaaaagcagttttctacaaatcgaactctccccaatc | 630 |
| Db | 301 | gacgtctatnacttagcccaaaagcagttttctacaaatnagacactttcccttaaatc | 360 |
| Oy | 631 | tgagacatggtgcgtttctcttaatgaaatgcagctccgcagtcagaattcatalatgtctcc | 690 |
| Db | 361 | tgagacatggtgcgtttctcttaatgaaatgcagctccgcagtcagaattcatalatgtctcc | 420 |
| Oy | 691 | acgtgcacagagtttaatgactctctgggtctctctatccagtttgctacccttggtgat | 750 |
| Db | 421 | actgtccacagagtttaatgactctctgggtctctctatccagtttgctacccttggtgat | 480 |
| Oy | 751 | ttgcgaactcttactatagaccatttgcctcggaagaaatttggtttcaataaatg | 810 |
| Db | 481 | ttgcgaactcttactatagaccatttgcctcggaagaaatttggtttcaataaatg | 540 |
| Oy | 811 | accgtccaaataatgatacacaagataagatacaagagtcgaaacttggtcatgctatgct | 870 |
| Db | 541 | accgtccaaataatgatacacaagataagatacaagagtcgaaacttggtcatgctatgct | 600 |
| Oy | 871 | ggcccaactcttaccctgtctcttaagaactcttgataaaatccctactgattcccttg | 930 |
| Db | 601 | ggcccaactcttaccctgtctcttaagaactcttgataaaatccctactgattcccttg | 660 |
| Oy | 931 | tacatctggtctctctgttgcgttgctgcacaaactctgacagatgaacaaatctccagcttc | 990 |
| Db | 661 | tacatctggtctctctgttgcgttgctgcacaaactctgacagatgaacaaatctccagcttc | 720 |
| Oy | 991 | atcatgcccattgttcaataaagtacactcactcttggaagacgtgaaactgaaaaactat | 1050 |
| Db | 721 | atcatgcccattgttcaataaagtacactcactcttggaagacgtgaaactgaaaaactat | 780 |
| Oy | 1051 | gaaagtttgccgatatagatgtggtgtccctctagataagaattttgtcaatgtagcgctca | 1110 |
| Db | 781 | gaaagtttgccgatatagatgtggtgtccctctagataagaattttgtcaatgtagcgctca | 840 |
| Oy | 1111 | aaaagatcttctatccaagaacgatatcttcacaggttgccactcactccaagaagaatt | 1170 |
| Db | 841 | aaaagatcttctatccaagaacgatatcttcacaggttgccactcactccaagaagaatt | 900 |
| Oy | 1171 | gtttgttcgacacttagtgaacagiatatctcactgatgaaatcagtcgttttggcc | 1230 |
| Db | 901 | gtttgttcgacacttagtgaacagiatatctcactgatgaaatcagtcgttttggcc | 960 |
| Oy | 1231 | catgaatctgctacatggtgcacaaaaacacatcgttaatatgtaattgtaactcttagtcaatg | 1290 |
| Db | 961 | catgaatctgctacatggtgcacaaaaacacatcgttaatatgtaattgtaactcttagtcaatg | 1020 |
| Oy | 1291 | caacacctctctcatcttctccctcttcacacagatctacagaatacactcatcttacac | 1350 |
| Db | 1021 | caacacctctctcatcttctccctcttcacacagatctacagaatacactcatctttacac | 1080 |
| Oy | 1351 | accttggtcttttcttaagaagaatcgaactggtgcagtttgttgatcccggttatcaag | 1410 |
| Db | 1081 | accttggtcttttcttaagaagaatcgaactggtgcagtttgttgatcccggttatcaag | 1140 |
| Oy | 1411 | gaattccacatcatcatgattatgtaatttaacagactatctaactccactcgaaagt | 1470 |

THIS PAGE BLANK (3PT0)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 12, 2002, 09:26:44 ; Search time 17.02 Seconds

(without alignments)
975.862 Million cell updates/sec

Title: US-09-165-460A-2

Perfect score: 2342
Sequence: 1 MEDUKTLDHPNIPMKLILS.....HPTLAERSTALDVSEKKKN 453

Scoring table: BLOSOM62
Gap 10.0, Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database: SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|--------------|---------------------|
| 1 | 2336 | 99.7 | 453 | 1 ST24_YEAST | P47154 saccharomyc |
| 2 | 903 | 38.6 | 474 | 1 ST24_SCHPO | Q10071 schizosacch |
| 3 | 698.5 | 29.8 | 475 | 1 FAC1_HUMAN | O75844 homo sapien |
| 4 | 328 | 14.0 | 426 | 1 YHFN_BACSU | P40769 bacillus su |
| 5 | 161 | 6.9 | 286 | 1 HTPX_PASMU | P57846 pasteurella |
| 6 | 150.5 | 6.4 | 310 | 1 HTPX_HELPJ | O92K84 helicobacte |
| 7 | 150 | 6.4 | 335 | 1 HTPX_ARCFU | O30004 archaeoglob |
| 8 | 147.5 | 6.3 | 283 | 1 HTPX_HAEIN | P44840 haemophilus |
| 9 | 142.5 | 6.1 | 310 | 1 HTPX_HELPY | O25582 helicobacte |
| 10 | 139 | 5.9 | 366 | 1 HTPX_AERPE | O9Y467 aeropyrum p |
| 11 | 138.5 | 5.9 | 258 | 1 HTPX_METTH | O26669 methanobact |
| 12 | 130.5 | 5.6 | 292 | 1 HTPX_PYRHO | O58997 pyrococcus |
| 13 | 130.5 | 5.6 | 317 | 1 HTPX_TREAC | O091V2 thermoplasma |
| 14 | 127.5 | 5.4 | 2136 | 1 YCF2_MARPO | P09975 marchantia |
| 15 | 120 | 5.1 | 292 | 1 HTPX_BUCAL | P57406 buchneria ap |
| 16 | 119.5 | 5.1 | 2294 | 1 YCF2_ARATH | P56786 arabidopsis |
| 17 | 119 | 5.1 | 297 | 1 HTPX_STRGC | O30795 streptococc |
| 18 | 116 | 5.0 | 293 | 1 HTPX_ECOLI | P23894 escherichia |
| 19 | 113 | 4.8 | 284 | 1 HTPX_METJA | O59076 methanococc |
| 20 | 113 | 4.8 | 298 | 1 HTPX_BACSU | O31657 bacillus su |
| 21 | 109 | 4.7 | 506 | 1 YCF2_EUGGR | P31921 europa gra |
| 22 | 109 | 4.7 | 595 | 1 YCF6_METJA | O58682 methanococc |
| 23 | 107.5 | 4.6 | 1032 | 1 MT18_YEAST | P40469 saccharomyc |
| 24 | 106.5 | 4.5 | 646 | 1 NTP1_HAEPV | O37319 heliothis a |
| 25 | 106.5 | 4.5 | 1029 | 1 YCF3_YEAST | P43571 saccharomyc |
| 26 | 105.5 | 4.5 | 873 | 1 COX1_ACACA | O37370 acanthamoeb |
| 27 | 105 | 4.5 | 1427 | 1 SRB8_YEAST | P25648 saccharomyc |
| 28 | 104.5 | 4.5 | 565 | 1 YHDS_YEAST | P38732 saccharomyc |
| 29 | 104.5 | 4.5 | 1100 | 1 LOS1_YEAST | P33148 saccharomyc |
| 30 | 104 | 4.4 | 302 | 1 HTPX_AOUAE | O67798 aquifex aeo |
| 31 | 104 | 4.4 | 530 | 1 CP14_CHICK | P79760 gallus gall |
| 32 | 103.5 | 4.4 | 185 | 1 GDU1_RAT | P35894 rattus norv |
| 33 | 103.5 | 4.4 | 450 | 1 VATH_SCHPO | O14265 schizosacch |

ALIGNMENTS

| | | | | | | |
|----|-------|-----|------|---|------------|--------------------|
| 34 | 103 | 4.4 | 645 | 1 | VP74_NPVAC | P15963 autographa |
| 35 | 103 | 4.4 | 1101 | 1 | KOE5_YEAST | 008217 saccharomyc |
| 36 | 103 | 4.4 | 577 | 1 | THM1_SCHPO | 009684 schizosacch |
| 37 | 102 | 4.4 | 2493 | 1 | YB46_YEAST | P35194 saccharomyc |
| 38 | 100.5 | 4.3 | 260 | 1 | YM16_RECAM | Q12466 reclinomona |
| 39 | 100.5 | 4.3 | 1240 | 1 | R1K1_SCHPO | Q10426 schizosacch |
| 40 | 100 | 4.3 | 1325 | 1 | MRP4_HUMAN | 015439 homo sapien |
| 41 | 99.5 | 4.2 | 309 | 1 | OT7A_HUMAN | 076100 homo sapien |
| 42 | 99 | 4.2 | 485 | 1 | YC11_KLEPN | Q48457 klebsiella |
| 43 | 99 | 4.2 | 1024 | 1 | RPOB_PLAFA | P21591 plasmodium |
| 44 | 98.5 | 4.2 | 313 | 1 | NU2M_RISA | 099617 rhipicephal |
| 45 | 98.5 | 4.2 | 417 | 1 | 065A_DROME | P82982 drosophila |

ALIGNMENTS

| | | | | |
|------------|---|-----------|------|---------|
| RESULT | 1 | | | |
| ST24_YEAST | ST24_YEAST | STANDARD: | PRT: | 453 AA. |
| AC | P47154; | | | |
| DT | 01-FEB-1996 (Rel. 33, Created) | | | |
| DT | 01-FEB-1996 (Rel. 33, Last sequence update) | | | |
| DT | 20-AUG-2001 (Rel. 40, Last annotation update) | | | |
| DE | CAAX PRENYL PROTEASE 1 (EC 3.4.24.-) (PRENYL PROTEIN-SPECIFIC ENDOPEPTIDASE 1) (PPEP 1) (A-FACTOR CONVERTING ENZYME). | | | |
| GN | STE24 OR AF1 OR YJRL17W OR J2032. | | | |
| OS | Saccharomyces cerevisiae (Baker's yeast). | | | |
| OC | Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; | | | |
| OC | Saccharomycetales; Saccharomycetaceae; Saccharomyces. | | | |
| OX | NCBI_TaxID=4932; | | | |
| NP | [1] | | | |
| RA | SEQUENCE FROM N.A. | | | |
| RA | Rose M., Koetter P., Entian K.D.; | | | |
| RL | Submitted (SEP-1995) to the EMBL/GenBank/DBJ databases. | | | |
| RP | [2] | | | |
| RX | SEQUENCE FROM N.A., AND FUNCTION. | | | |
| RA | MEDLINE=97167681; PubMed=9015299; | | | |
| RT | Fujimura-Kamada K., Nouvet F.J., Michaelis S.; | | | |
| RT | "A novel membrane-associated metalloprotease, Ste24p, is required for the first step of NH2-terminal processing of the yeast a-factor precursor." | | | |
| RL | J. Cell Biol. 136:271-285(1997). | | | |
| RP | [3] | | | |
| RX | FUNCTION. | | | |
| RA | MEDLINE=97218305; PubMed=9065405; | | | |
| RT | Boyarichuk V.L., Ashby M.N., Rine J.; | | | |
| RT | "Modulation of Ras and a-factor function by carboxyl-terminal proteolysis." | | | |
| RL | Science 275:1796-1800(1997). | | | |
| RP | [4] | | | |
| RX | FUNCTION. | | | |
| RA | MEDLINE=98365461; PubMed=9700155; | | | |
| RT | Tam A., Nouvet F.J., Fujimura-Kamada K., Slunt H., Sisodia S.S., | | | |
| RT | Michaelis S.; | | | |
| RT | "Dual roles for Ste24p in yeast a-factor maturation: NH2-terminal proteolysis and COOH-terminal CAAX processing." | | | |
| RL | J. Cell Biol. 142:635-649(1998). | | | |
| RP | [5] | | | |
| RX | FUNCTION. | | | |
| RA | MEDLINE=98393572; PubMed=9725832; | | | |
| RT | Boyarichuk V.L., Rine J.; | | | |
| RT | "Roles of prenyl protein proteases in maturation of Saccharomyces cerevisiae a-factor." | | | |
| RL | Genetics 150:95-101(1998). | | | |
| RP | [6] | | | |
| RX | FUNCTION. | | | |
| RA | MEDLINE=20158928; PubMed=10692417; | | | |
| RT | Schmidt W.K., Tam A., Michaelis S.; | | | |
| RT | "Reconstitution of the Ste24p-dependent N-terminal proteolytic step in yeast a-factor biogenesis." | | | |
| RL | J. Biol. Chem. 275:6227-6233(2000). | | | |
| RP | [7] | | | |

CHARACTERIZATION.
MEDLINE-20285442; PubMed-10825201;
RA Trublood C.E., Boyartchuk V.L., Pictologlou E.A., Rozema D.,
RA Poulter C.D., Rine J.,
RT "The Caax proteases, Aiclp and Rcelp, have overlapping but distinct
RT substrate specificities."
RL Mol. Cell. Biol. 20:4381-4392(2000).
RN [8]
RP SUBCELLULAR LOCATION.
RX MEDLINE-98409630; PubMed-9736709;
RA Schmidt W.K., Tam A., Fujimura-Kanada K., Michaelis S.,
RT "Endoplasmic reticulum membrane localization of Rcelp and Ste24p,
RT yeast proteases involved in carboxyl-terminal CAAX protein processing
RT and amino-terminal a-factor cleavage."
RL Proc. Natl. Acad. Sci. U.S.A. 95:11175-11180(1998).
CC -1- FUNCTION: PROTEOLYTICALLY REMOVES THE C-TERMINAL THREE RESIDUES OF
CC FARNESYLATED A-FACTOR MATING PHEROMONE. ALSO ACTS TO CLEAVE THE N-
CC TERMINAL EXTENSION OF THE PHEROMONE. DOES NOT ACT ON RAS.
CC -1- COFACTOR: BINDS ONE ZINC ION (PROBABLE).
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. ENDOPLASMIC
CC RETICULUM.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M48 (ZINC
CC METALLOPROTEASE).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).

DR EMBL: 249617; CAAB9647.1;
DR EMBL: 077137; AAB38271.1;
DR MEROPS: M48.001;
DR SGD: S0003878; STE24.
DR InterPro: IPR001915; Peptidase_M48.
DR InterPro: IPR000130; Zn_MTPeptide.
DR Pfam: PF01435; Peptidase_M48; 1.
DR PROSITE: PS00142; ZINC_PROTEASE; 1.
DR Hydrolase: Metalloprotease; Zinc; Transmembrane;
KW Endoplasmic reticulum; Pheromone response.
FT TRANSMEM 13
FT TRANSMEM 33
FT TRANSMEM 90
FT TRANSMEM 122
FT TRANSMEM 142
FT TRANSMEM 168
FT TRANSMEM 188
FT TRANSMEM 198
FT TRANSMEM 218
FT TRANSMEM 307
FT TRANSMEM 327
FT TRANSMEM 358
FT TRANSMEM 378
FT METAL 257
FT ACT_SITE 298
FT METAL 301
FT METAL 301
FT METAL 390
FT ACT_SITE 394
FT SEQUENCE 453 AA; 52324 MW; 33JC6AE2D7C99DA CRC64;

Query Match 99.7%; Score 2336; DB 1; Length 453;
Best Local Similarity 99.8%; Pred. No. 3.7e-160;
Matches 452; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MEDLKITLHPNIPMKLIISGSIASFSESYTYVOYOKLSTKLPVLEDEIDETPH 60
DB 1 MEDLKITLHPNIPMKLIISGSIASFSESYTYVOYOKLSTKLPVLEDEIDETPH 60
QY 61 KSNVSRARAKSIFEGDVNLAOKLVFIKYDLPKRTWMAVSLNVLVPRFMSTVAQ 120
DB 61 KSNVSRARAKSIFEGDVNLAOKLVFIKYDLPKRTWMAVSLNVLVPRFMSTVAQ 120
QY 121 SCLFGLSSLTLDVLPSTYSYSHFVLEKFGFNKLTVOIMIDMKSLTLVAIGPIL 180
DB 121 SCLFGLSSLTLDVLPSTYSYSHFVLEKFGFNKLTVOIMIDMKSLTLVAIGPIL 180

QY 181 YLFKIFDKFPDPLMYINVFLEVVQILAMTIIPVIMPMFKFTPLEDEGLKSTESLA 240
DB 181 YLFKIFDKFPDPLMYINVFLEVVQILAMTIIPVIMPMFKFTPLEDEGLKSTESLA 240
QY 241 DRVGFPDLPKIFVYIDSKSSSHSNAYFTGLPFTSKRIVLFDLTVNSNSTDEIRAVLAHEIG 300
DB 241 DRVGFPDLPKIFVYIDSKSSSHSNAYFTGLPFTSKRIVLFDLTVNSNSTDEIRAVLAHEIG 300
QY 301 HKQKHVYVMVTFSQLHFLIFSLTSTYRNTSFYNTFGFPLEKSTGTFVDPVITKEPPI 360
DB 301 HKQKHVYVMVTFSQLHFLIFSLTSTYRNTSFYNTFGFPLEKSTGTFVDPVITKEPPI 360
QY 361 IIGFMLFNDLTPLECAMQFVMSLSTRHEVQADAVAKKLGKONLCRALIDLOIKNST 420
DB 361 IIGFMLFNDLTPLECAMQFVMSLSTRHEVQADAVAKKLGKONLCRALIDLOIKNST 420
QY 421 MNVDPDYSSHYSHPTLALRSTALDYSEKKKN 453
DB 421 MNVDPDYSSHYSHPTLALRSTALDYSEKKKN 453

RESULT 2
ST24_SCHPO STANDARD: PRT: 474 AA.
ID ST24_SCHPO
AC Q10071.
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE PROBABLE CAAX PRENYL PROTEASE 1 (EC 3.4.24.-) (PRENYL PROTEIN-
DE SPECIFIC ENDOPEPTIDASE 1) (PPSEP 1).
GN SPAC3H1.05.
OS Schizosaccharomyces pombe (fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomyces.
OX NCBI_Taxid=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RA Gentles S., Churcher C.M., Barrell B.G., Rajandream M.A., Walsh S.V.,
RL Submitted (DEC-1995) to the EMBL/Genbank/DBJ databases.
CC -1- FUNCTION: PROTEOLYTICALLY REMOVES THE C-TERMINAL THREE RESIDUES OF
CC FARNESYLATED PROTEINS (BY SIMILARITY).
CC -1- COFACTOR: BINDS ONE ZINC ION (POTENTIAL).
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. ENDOPLASMIC
CC RETICULUM (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M48 (ZINC
CC METALLOPROTEASE).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).

DR EMBL: 268144; CAAB92258.1;
DR MEROPS: M48.001;
DR InterPro: IPR001915; Peptidase_M48.
DR InterPro: IPR000130; Zn_MTPeptide.
DR Pfam: PF01435; Peptidase_M48; 1.
DR PROSITE: PS00142; ZINC_PROTEASE; 1.
KW Hypothetical protein; Hydrolase; Metalloprotease; Zinc; Transmembrane;
KW Endoplasmic reticulum.
FT TRANSMEM 103
FT TRANSMEM 123
FT TRANSMEM 196
FT TRANSMEM 216
FT TRANSMEM 230
FT TRANSMEM 250
FT TRANSMEM 344
FT TRANSMEM 364
FT TRANSMEM 381
FT TRANSMEM 401
FT METAL 332
FT METAL 332
FT ACT_SITE 333
FT METAL 336
FT METAL 336
FT SEQUENCE 453 AA; 52324 MW; 33JC6AE2D7C99DA CRC64;

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 12, 2002, 13:20:11 ; Search time 4622.84 Seconds

(Without alignments)
10520.305 Million cell updates/sec

Title: US-09-165-460a-3
Perfect score: 2948

Sequence: 1 tgaactgttgatgacacaag.....ggggagagataaagaatcaca 2948

Scoring table: IDENTITY_NTC
Gapop 10.0 , Gapept 1.0

Searched: 1472140 segs, 8248589755 residues

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seg length: 0
Maximum DB seg length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database:

GenDb1: *
1: gb_da: *
2: gb_htg: *
3: gb_in: *
4: gb_om: *
5: gb_ov: *
6: gb_pat: *
7: gb_ph: *
8: gb_pl: *
9: gb_pr: *
10: gb_ro: *
11: gb_sts: *
12: gb_sy: *
13: gb_un: *
14: gb_vl: *
15: em_da: *
16: em_fun: *
17: em_hum: *
18: em_in: *
19: em_om: *
20: em_or: *
21: em_ov: *
22: em_pat: *
23: em_ph: *
24: em_pl: *
25: em_ro: *
26: em_sts: *
27: em_sy: *
28: em_un: *
29: em_vl: *
30: em_htgo_hum: *
31: em_htgo_inv: *
32: em_htgo_rod: *
33: em_htg_hum: *
34: em_htg_inv: *
35: em_htg_rod: *
36: em_htg_other: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB | ID | Description |
|------------|--------|-------------|--------|----|------------|----------------------|
| 1 | 2948 | 100.0 | 29987 | 8 | SC8156 | 249260 S.cerevisiae |
| 2 | 1560.4 | 52.9 | 4156 | 8 | D50276 | D50276 Saccharomyces |
| 3 | 1361 | 46.2 | 4130 | 8 | YSCNCRIG | 147821 Saccharomyces |
| 4 | 1163 | 39.5 | 4225 | 8 | SCU5682 | 065682 Saccharomyces |
| 5 | 1069.4 | 36.3 | 3250 | 8 | SCU32380 | 073580 Saccharomyces |
| 6 | 931 | 31.6 | 3525 | 8 | SCU63849 | 063849 Saccharomyces |
| 7 | 786.8 | 26.7 | 3240 | 8 | YSCBUTL1 | D50083 Yeast BUTL |
| 8 | 703.8 | 23.9 | 3020 | 8 | SCRDSTGEN | X88991 S.cerevisia |
| 9 | 725.2 | 7.6 | 270 | 8 | YSCGRC5X | M87379 Yeast Eco R |
| 10 | 143.8 | 4.9 | 27559 | 8 | SC8338 | Z43210 S.cerevisia |
| 11 | 75 | 2.5 | 2023 | 8 | SCYNR68C | Z71683 S.cerevisia |
| 12 | 71.8 | 2.4 | 4775 | 8 | SCU32938 | U32938 Saccharomyces |
| 13 | 68 | 2.3 | 780 | 11 | CNS06K8B | AL402561 T7 end of |
| 14 | 65.6 | 2.2 | 256172 | 2 | AC005139 | AC005139 Plasmodiu |
| 15 | 65.6 | 2.2 | 310779 | 2 | AC005140 | AC005140 Plasmodiu |
| 16 | 64.4 | 2.2 | 904 | 11 | CNS06I89 | AL39967 T3 end of |
| 17 | 64.4 | 2.2 | 86827 | 3 | PFMAL3P5 | AL034556 Plasmodiu |
| 18 | 63.4 | 2.2 | 1920 | 3 | PGRTB | X62393 P.falciparu |
| 19 | 63.4 | 2.2 | 110000 | 2 | PFMAL3P2_0 | AL049185 Plasmodiu |
| 20 | 62 | 2.1 | 253305 | 3 | PFMAL3P7 | AL034559 Plasmodiu |
| 21 | 61.6 | 2.1 | 11441 | 6 | AX083744 | AX083744 Sequence |
| 22 | 60.8 | 2.1 | 234112 | 3 | PFMAL4P2 | AL035475 Plasmodiu |
| 23 | 60.6 | 2.1 | 202872 | 2 | AC016160 | AC016160 Homo sapl |
| 24 | 60 | 2.0 | 5632 | 3 | AF310892 | AF310892 Dictyoste |
| 25 | 59.4 | 2.0 | 3214 | 3 | AF019980 | AF019980 Dictyoste |
| 26 | 59.4 | 2.0 | 318221 | 3 | PFMAL3P3 | AL049184 Plasmodiu |
| 27 | 58.8 | 2.0 | 321003 | 2 | PFMAL4P3 | AL035476 Plasmodiu |
| 28 | 58.4 | 2.0 | 78439 | 17 | HSU85197 | U85197 Homo saplen |
| 29 | 58.4 | 2.0 | 169546 | 2 | AC004157 | AC004157 Plasmodiu |
| 30 | 58.4 | 2.0 | 245802 | 2 | AC006279 | AC006279 Plasmodiu |
| 31 | 58.4 | 2.0 | 251124 | 3 | HUAE000660 | AE000660 Homo sapl |
| 32 | 58.2 | 2.0 | 153098 | 3 | PFMAL3P2 | AL034558 Plasmodiu |
| 33 | 58 | 2.0 | 169794 | 2 | AC004688 | AC004688 Plasmodiu |
| 34 | 57.8 | 2.0 | 178137 | 9 | AC011302 | AC011302 Homo sapl |
| 35 | 57.8 | 2.0 | 13684 | 3 | AE001403 | AE001403 Plasmodiu |
| 36 | 57.8 | 2.0 | 153477 | 2 | AC006478 | AC006478 Plasmodiu |
| 37 | 57.6 | 2.0 | 53932 | 2 | AC023371 | AC023371 Homo sapl |
| 38 | 57 | 1.9 | 12029 | 3 | AE001431 | AE001431 Plasmodiu |
| 39 | 57 | 1.9 | 163443 | 2 | AC006280 | AC006280 Plasmodiu |
| 40 | 57 | 1.9 | 178273 | 2 | AC005308 | AC005308 Plasmodiu |
| 41 | 57 | 1.9 | 196149 | 2 | AC004709 | AC004709 Plasmodiu |
| 42 | 56.8 | 1.9 | 67970 | 3 | PFMAL1P3 | AL031746 Plasmodiu |
| 43 | 56.8 | 1.9 | 149627 | 9 | AC087428 | AC087428 Homo sapl |
| 44 | 56.6 | 1.9 | 92633 | 2 | PFMAL4P1_3 | Continuation (4 of |
| 45 | 56.6 | 1.9 | 153477 | 2 | AC006278 | AC006278 Plasmodiu |

ALIGNMENTS

RESULT 1
SC8156/c
LOCUS SC8156 29987 bp DNA
DEFINITION S.cerevisiae chromosome XIII cosmid 8156.
ACCESSION Z49260 Z71257
VERSION Z49260.1 GI:809081
KEYWORDS BUTL; cytochrome b5; DAG1; Initiation factor 1A; Inorganic pyrophosphatase; Ipp2; nitrate reductase; Orotate phosphoribosyltransferase; tau element; TIF1; TP53; transfer RNA-Gln; trehalose-phosphate synthase; U6 snRNP; URA10. Baker's yeast.
SOURCE Saccharomyces cerevisiae
ORGANISM Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; Saccharomycetales; Saccharomycetaceae; Saccharomyces.
REFERENCE 1 (bases 1 to 29987)
AUTHORS Live-G. and Churcher,C.M.
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 29987)
AUTHORS Bartell,B. and Rajandream,M.A.
TITLE Direct Submission

THIS PAGE BLANK (1-3870)

JOURNAL

COMMENT

Submitted (12-MAY-1995) *Saccharomyces cerevisiae* chromosome XIII
sequencing project, Sanger Centre, Hinxton Hall, Hinxton, Cambridge
CB10 1RQ. E-mail: barrel@sanger.ac.uk

Notes:

All CDS over 100 codons have been analysed. CDS that are completely
overlapped and those that are overlapped by more than 50%
of their length by a larger CDS have been omitted from this
analysis. Details of the omitted CDS are available on request. The more
significant matches with motifs in the PROSITE database are
also included but some of these may be fortuitous. The length in
codons and the calculated codon adaptation index (CAI)
is given for each CDS.

Cosmid 8156 is overlapped at the start by cosmid 9920, emb1 entry
SC9920,
accession no. Z48639 and at the end by cosmid 8021.
location/Qualifiers

1. 29987
/organism="Saccharomyces cerevisiae"
/strain="AB972"
/db_xref="taxon:4932"
/chromosome="XIII"
/clone="cosmid 8156"
/map="13R"
complement(<1..356)
/note="YMB156.01c, unknown, len: 118, CAI: 0.14, overlaps
and extends YW9920.13c"
/codon_start=1
/product="unknown"
/protein_id="CA89242.1"
/db_xref="GI:809082"
/db_xref="SWISS-PROT:Q03496"
/translation="MTDVQSSQIEELFRYKELIANPKNINENKDAVLQIEHD
FRYIYIKDGLPMLNESTRILPDPFSICILRSHQIASKIDSQELFLAVKEQLLT
EANNIIEFYLDLFWA"
1. 140
/note="overlap with SC9920 Z48639 S. cerevisiae chromosome
XIII cosmid 9920"
complement(851..1312)
/gene="TIF11"
complement(851..1312)
/note="TIF11"
/gene="TIF11"
/note="YMB156.02c TIF11 gene, IFLA_YEAST, P3891,
initiation factor 1a"
/codon_start=1
/product="Tif1p"
/protein_id="CA89243.1"
/db_xref="GI:809083"
/db_xref="SWISS-PROT:P38912"
/translation="MGKNTKGGKRGKRGKNDSDGPKRELLYKEGGEYAOITKLN
GRVASCDFGNKRNHIRGLKTKYVMGQDIIIVSLDPODDQDVYHKNLDEART
LKNQGLPENAKINETDNFGESDEDFVNEFGNDEDEDEDEDELDIDI"
complement(1678..4842)
/gene="TPS3"
complement(1678..4842)
/note="TPS3"
/note="YMB156.03c, TPS3 gene, len: 1054, CAI: 0.21,
SW:TPS3_YEAST P38426, alpha, alpha-trehalose-phosphate
synthase and PIR:S27471, PS00213 Lipocalin signature;
translation in TPS3_YEAST starts at amino acid 33 in this
sequence"
/codon_start=1
/product="Tps3p"
/protein_id="CA89244.1"
/db_xref="GI:809084"
/db_xref="SWISS-PROT:P38426"
/translation="MTIIVASLEFLPYTPOFEADVNSDTAKLVESMIRKVDNNOELS
NNKORSSSVTSASSHYIGLPOEADINGPPLRANVGPATGVNHNEMELSSFOFL
BELTANATHAANGIPRANPYSSSGSTARQREVEFFSAPSRVCSPOEASASISIA
SRSSAHNDLSSILKKNPVLSTDSHPPRVSSSKSAVITPVSKSVDPDPAVDVAKY
REEFOQASLPSMKRVSGSTAGDSIASSSMLRISQOFIDTIDTSDEDDIDLE
TDAIKYVPEKGGYSNNAKLRASISLRASIELEKLPITIVSDRSGNSLKNAAVIAV
AEKTVKEPVSWGTGIPIDELPHEVCHRSIKKLEDDFSFPVTDITFGAYKXNYA

FEATURES

source

CDS

misc_feature

gene

CDS

CDS

gene

1. 140
/note="overlap with SC9920 Z48639 S. cerevisiae chromosome
XIII cosmid 9920"
complement(851..1312)
/gene="TIF11"
complement(851..1312)
/note="TIF11"
/gene="TIF11"
/note="YMB156.02c TIF11 gene, IFLA_YEAST, P3891,
initiation factor 1a"
/codon_start=1
/product="Tif1p"
/protein_id="CA89243.1"
/db_xref="GI:809083"
/db_xref="SWISS-PROT:P38912"
/translation="MGKNTKGGKRGKRGKNDSDGPKRELLYKEGGEYAOITKLN
GRVASCDFGNKRNHIRGLKTKYVMGQDIIIVSLDPODDQDVYHKNLDEART
LKNQGLPENAKINETDNFGESDEDFVNEFGNDEDEDEDEDELDIDI"
complement(1678..4842)
/gene="TPS3"
complement(1678..4842)
/note="TPS3"
/note="YMB156.03c, TPS3 gene, len: 1054, CAI: 0.21,
SW:TPS3_YEAST P38426, alpha, alpha-trehalose-phosphate
synthase and PIR:S27471, PS00213 Lipocalin signature;
translation in TPS3_YEAST starts at amino acid 33 in this
sequence"
/codon_start=1
/product="Tps3p"
/protein_id="CA89244.1"
/db_xref="GI:809084"
/db_xref="SWISS-PROT:P38426"
/translation="MTIIVASLEFLPYTPOFEADVNSDTAKLVESMIRKVDNNOELS
NNKORSSSVTSASSHYIGLPOEADINGPPLRANVGPATGVNHNEMELSSFOFL
BELTANATHAANGIPRANPYSSSGSTARQREVEFFSAPSRVCSPOEASASISIA
SRSSAHNDLSSILKKNPVLSTDSHPPRVSSSKSAVITPVSKSVDPDPAVDVAKY
REEFOQASLPSMKRVSGSTAGDSIASSSMLRISQOFIDTIDTSDEDDIDLE
TDAIKYVPEKGGYSNNAKLRASISLRASIELEKLPITIVSDRSGNSLKNAAVIAV
AEKTVKEPVSWGTGIPIDELPHEVCHRSIKKLEDDFSFPVTDITFGAYKXNYA

misc_feature

CDS

CDS

CDS

CDS

CDS

KOLIMPTLHIOIPDNPNSKAFEDHSDMYTQRYNOKRSDRISVYKPGDITWIDHILM
LVPQVRRKLPRADKIGFLVHSFSEFCEFRANERILEGIGANVFOQRKIRK
FLQCNRLKRVNDSEVYKHNIVSVYAPIGIYVHLISQLRNSVYRQLIER
WRNKLTICRQDFORIGLOKMLAYEELIENPPIEYVILQICIGSDPEYERQ
IIVVDRINSSNISISOPVYVFLHODIFAOYLILNCADVPLDALBGMATYCHE
FIVSFEENAPLLSEFYGSSSVLKEGAILNPWQINHAQSIKRLSEAPKRRRW
KLFEVYEDSDMTIKFCFENINNAMSNCTSTYFNLAPEKFCADYKASKEHLIF
KISEPTEBRMLSLSELSNNIVYLSPTKPSLNGVYINGLIAGNAYVAVNG
SWINAYEELDMKKEVAKIFDERVEELPGSYXYIADSMTEPHENDDOORVPVIGEA
ITHIVLDDNDIHKVADIVFOQTGLALAAEFLAKFYKISGVSPTNRSLSRQ
SSMSVGNKKRHPQVQVDFCVSGSTPILEFLYKQVEKRNKLPYITLYGSSR
STYAKHEHNGVNEFLTIHDLTA"
complement(2260..2301)
/gene="TPS3"
/note="PS00213 Lipocalin signature"
5199..6140
/note="YMB156.04, unknown, len: 313, CAI: 0.11, previously
noted ORF, SW:YTP3_YEAST P38430"
/codon_start=1
/product="unknown"
/protein_id="CA89245.1"
/db_xref="GI:809085"
/db_xref="SWISS-PROT:P38430"
/translation="NKKLYDAHCHVITDPEDNTEFCGDDGSGQGLRCVMSNPYDRNL
KLIAGRSKNDICVGFVHPWYSHLFYVGSBRDVSHTQDLYEKNESQFSLVOYL
PEPDLIEYIKREFNDLYSYIGEIGDLFLPLNANGFMONEKALTYKYTLSDQE
TVRFPRFLARFTSKPSIHDVCHKGLNDICNEELIYHSVKICLHSTGSKETLIG
QWLKRFPPRIFVSLSKWINEFDEGDLALPSTCLITETDYPIDNPPSYQAL
TEOLOYLNAQIARADELIDASQALRAYENEOKFIK"
6392..6997
/note="YMB156.05, unknown, len: 301, CAI: 0.12,
SW:YTP2_YEAST P38429 previously noted ORF"
/codon_start=1
/product="unknown"
/protein_id="CA89246.1"
/db_xref="GI:809086"
/db_xref="SWISS-PROT:P38429"
/translation="NARVYNINAEISRSRGRPTGGGYASNNNSCINNNSNNNNNN
NNNNNSNNNSNNNGPTSSGRNGKQRLPAQOQYIKNLITHTDHPDLRPSH
DEEYTDARFLARKYHFDLPDNLGLGILGASAKTYSYKRNQGGHDKRIHR
DLANVRRHFEHSIKETDICIPOFIYKYNKKKFKKEFG"
7278..7889
/note="YMB156.06, unknown, len: 301, CAI: 0.13,
SW:YTP1_YEAST P38428, extends previously noted ORF"
/codon_start=1
/product="unknown"
/protein_id="CA89247.1"
/db_xref="GI:809087"
/db_xref="SWISS-PROT:P38428"
/translation="MEDSRLLITLILVGVITLKTFFQSNOHPSAQLSATGVNAGR
PGOSTONALARTGVNGHPTVQWETVONLAPNLPEDQIRISLEYSVEETVERT
LRGDESFEPGEPSPRABGANAAVNNAGGGEFNDPRKKNMICAEYNLIDKFEVDLN
EDMSNLSFDOLDIEERKRLIMQARKNLEKLSQDNQVSLT"
complement(8013..9398)
/note="YMB156.07c, unknown, len: 461, CAI: 0.11"
/codon_start=1
/product="unknown"
/protein_id="CA89248.1"
/db_xref="GI:809088"
/db_xref="SWISS-PROT:Q03508"
/translation="MEFEFERRRKRGSSNCGNRYLAKNNSCACELAPEVEBEHAI
TKLHIDFQNTLFPATPGTLPOLYRELNLITLSTLNGGMNPGFLQALIEISKIK
PRRYSNMADIVKLAESEYSKADITISVLGRSESKFLIEHALQTAIRKSCSENER
RBNAYCVKRAISIEYTSKYKELMRDLLEYVPSLRISTYDDRHOIDAFKSPFSLD
LPRKNSAIVPRPFTKSLPREQELKAYMDVARKNNSGALSTSQGFDRRTROJGYIL
CTASHRLSLLEVTKYLRKRGRTFRKTLKHPYLAIPASAGKQIPALELAKVNSND
TTFDSKRTKQVHISQITLYLQDQPCIVHFOVTDLAVYCAHGNKRLPLEVYKATPEP
NRTFLTEPEYIYTGHYKRDRIEDLEVTERLINCEDEDIHVPLDWTIPKAFGRF
AKLAAPICSA"
9991..12852
/note="YMB156.08, unknown, len: 953, CAI: 0.20, probable
integral membrane glycoprotein"
/codon_start=1

THIS PAGE BLANK

```

/note="Xenopus laevis, 09, 14P2 gene (PPA2), len: 310, CBI: 0.14,
/SW: PFY2-EXAST_P28239, inorganic pyrophosphatase,
/mitochondrial, contains P500387 Inorganic pyrophosphatase
signature"
/codon_start=1
/product="Ipp3p"
/protein_id="C0A89250.1"
/db_xref="GI:809090"
/db_xref="SWISS-PROT:P28239"
/gene_synonym="MILNNAALPITSKARSTERLKQPTNLISIRNHRQSTIQGSKYTT
TGNKATLLNGEVSFHFQPDLDLNEHETVMATVEVPMTGTFEISKELENNPITV

```

| | | | |
|----|-------|---|-------|
| QY | 481 | caaaagacacctaataatggaacttggaaacttaataagaacacttaaggaacttaattcc | 540 |
| Db | 27305 | CAAAAGACACTTAATATGGAACCTTGGAAATTAATAGGACATTAGGAACCTTAATTC | 27248 |
| QY | 541 | aagtttgaagccgacctatgttgaagtttattcgcttcgctcgaagtaaatgaatttga | 600 |
| Db | 27245 | AAGTTTGAAGCCGACCTATGTTGAGGTTTATTCGCTTGAGTAATGATTAAATTGA | 27166 |
| QY | 601 | aaacacatttggcgttagcgaagaattgataccctaatttcgtttaagcaagtgcacaaat | 660 |
| Db | 27185 | AAACCATTTGTGGCGTAGGGAAGATGTGATATCCCATTTCTGTTAGGCAAGTCACAAATA | 27126 |
| QY | 661 | aaaaaaccttgaaaaaatttcgttctacttttcttatgatatagatatgatgatggtt | 720 |
| Db | 27125 | AAAAACCTTGGAAAAAATTTCTGTTACTTTCTTAATAGATAATATATATGATAGGTTT | 27066 |
| QY | 721 | gctataagatgaaggtattatcgcgtcccttcttgatcccatataataaatctctt | 780 |
| Db | 27065 | GCTTATAGATGAAGTATTTATTCGCGTCTTGTGATATCCCATTAATTAATAATCTTT | 27006 |
| QY | 781 | taaaatgcatttctcgtgtccttcttgcttcctgtatcttcttttttttttgagcaactg | 840 |
| Db | 27005 | TAAATGCATTTCTCGGCGCTTTTGCTTGTGCTCGATTTTCTTTTGTGACACACG | 26948 |
| QY | 841 | gatggaaaaccttggatgatattatacccttattttaagtaccacaaatcagagattc | 900 |
| Db | 26945 | GATGGAACCTTTGATGATATTTTATACCTTATTTTAAGTACTCAAAAATACGAGATT | 26886 |
| QY | 901 | caggacacaaacatagaattcttcttcttcgagaaaaaataaagcgaataaatttgact | 960 |
| Db | 26885 | CAGGACAAACATAGAAATTTCTTTTGCAAGAAAAATAAACGAATTAATATGATGCT | 26822 |
| QY | 961 | tgactactgactgtctgcataagagagacacagacgaatgcatactccaaact | 1020 |
| Db | 26825 | TGACTACTGACTGTCTGCATAGAGAGAACACGAACCAATGCTATTCATAACT | 26766 |
| QY | 1021 | tctagtctccatacatctccatactcctatgctctacgcgtatactgcaacttccaaac | 1080 |
| Db | 26765 | TCTAGTCTCCTATACACATCTCCATATCCTATGCTACCGTATATGCAACTTCAACAC | 26706 |
| QY | 1081 | agaagggctcctaagcagaatactcccgagagattaaactcgcagtcgacaaaacttaacat | 1140 |
| Db | 26705 | AGAAAGGCTTAAGAGAGATTAATCTCCGACATTAATATCGCATGCAAAAACCTTAACAT | 26648 |
| QY | 1141 | tatgctaaattccaaaccttttcttctgtgcctttttaaactcccatatctcagtaaac | 1200 |
| Db | 26645 | TATGCTAATTTCCAAACCTTTTGTGTCCTTTTAAACTCAATATCTAGTACACAC | 26586 |
| QY | 1201 | ttcacataataagttcgaagacgaacttttaagcttaagttatccagttatccagttacgc | 1260 |
| Db | 26585 | TTCACATTAATTAAGTTCCAAAGACGACATTTTGTGGCTTAAGTATTAATCCACAGTTATTAACG | 26528 |
| QY | 1261 | tgcaattgcacaaaccttgcgaacttcagcgcaagtctgtgaaagaaacttaagaaatgtgtgc | 1320 |
| Db | 26525 | TGCAATTGCCAAACCTTGGCAATTCAGCCACTTGTGTAAACCTTAAGAAATGTGTGC | 26466 |
| QY | 1321 | gagtttatgaccttataattgttgaccggtttatagattttglatatatacatttaataa | 1380 |
| Db | 26465 | GATGTTATTTGACCTTATATTGTGGAACCGCTTTAGATTGTTATTAATCAATTAATTA | 26406 |
| QY | 1381 | tccaaagagcctatactctgaagatttttccacgatctcctgaatacttggagttcag | 1440 |
| Db | 26405 | TCCAAAAGCGCTATACTTGAAGATTTTACACGAATTCCTGAAATATTTGGAGTTTCAG | 26346 |
| QY | 1441 | gaatttatatttgcacaaataaacttggagaaatttttaacgcgcaatgccttggactac | 1500 |
| Db | 26345 | GAAATTTATTTCCACCAATATCTGAGGAATATTTTACACGTCACGCTTTTGACATAC | 26286 |
| QY | 1501 | gtacttaaaccttaataccgacttcgcaactaagctatcaacaagtatttttggcaacatc | 1560 |
| Db | 26285 | GTACTTAACCTTAATACCGCACTTGGCAACTAAGCATATCAACAGTATTTTGGCAACATC | 26226 |
| QY | 1561 | gcttttttttggacttgcgacgcacacacactgcttatagagcaattacagaaggtccat | 1620 |

THIS PAGE BLANK (COPY)

| | | | | |
|----|-------|------------------------------------|------------------------------------|-------|
| Db | 25145 | CCGGTAGAGACCTTGCATGGGAGGGGTAAC | TCTAGAGGATAGAAATCTCCCAATTAAGACTTAA | 25086 |
| Oy | 2701 | caaggtactactactaggatatataaaagaagtg | tctactctccatccaggaaataatgtct | 2760 |
| Db | 25085 | CAAGGTACTACTTAGGTATATAAAAAAGGTT | CTCAGCTCCATCCAGGAATTTAAATGCT | 25026 |
| Oy | 2761 | aacgtattaagaagaaacattatattggtgtcc | gcgccaatacaacacccctaaagttaagct | 2820 |
| Db | 25025 | AACGTATTAAGAAGAAACCTTATATATGGGT | TCCCGCCCAATCAACACCCCTACAGTTTAAGCT | 24966 |
| Oy | 2821 | gataattccctagagctgtgtcaagatactact | ttaaaaaattacaacttaagcgacaatggt | 2880 |
| Db | 24965 | GATATTTCCTAGAGGCTTGTCACAAATCTCT | TACAAAATATACACTTAAGCCACATGCT | 24900 |
| Oy | 2881 | gaagctaattgttggaatagacaatgaataaata | cgaatctatctgagagataatgaggagataaa | 2940 |
| Db | 24905 | GAAGATTAATGATGGGAATACCATGAAAAAT | AACGATTTGAGCGATTAATGGGAGGATATAA | 24846 |
| Oy | 2941 | gaatcaca | 2948 | |
| Db | 24845 | GAAATCACA | 24838 | |

| RESULT 2 | LOCUS | DEFINITION | ACCESSION | VERSION | KEYWORDS | SOURCE | ORGANISM |
|----------|---|--|-----------|-------------|----------|--------|----------|
| D50276 | 4156 bp | DNA | PLN | 10-FEB-1999 | | | |
| D50276 | | Saccharomyces cerevisiae HST1 gene, complete cds. | | | | | |
| D50276 | | | | | | | |
| D50276.1 | GI:1777313 | HST1, high-copy suppressor of TFP sensitivity. | | | | | |
| | | Saccharomyces cerevisiae (strain:RAY-3a) DNA. | | | | | |
| | | Saccharomyces cerevisiae | | | | | |
| | | Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; | | | | | |
| | | Saccharomycetales; Saccharomycetaceae; Saccharomyces. | | | | | |
| | 1 (bases 1 to 4156) | | | | | | |
| | | Tsuchiya.E. | | | | | |
| | Direct Submission | | | | | | |
| | Submitted (13-APR-1995) | to the DDBJ/EMBL/GenBank databases. Eiko | | | | | |
| | Tsuchiya, Hiroshima University Faculty of Engineering, Department | | | | | | |
| | of Fermentation Technology; Kagami-Yama, Higashi-Hiroshima, | | | | | | |
| | Hiroshima 739, Japan (E-mail:etsuchi@ipc.hiroshima-u.ac.jp, | | | | | | |
| | Tel:0824-24-7868, Fax:0824-24-7868). | | | | | | |
| | 2 (bases 1 to 4156) | | | | | | |
| | Tsuchiya.E., Matsuzaki.G., Tsukao.A. and Miyakawa.T. | | | | | | |
| | Molecular cloning and characterization of S. cerevisiae HST1 gene | | | | | | |
| | Unpublished (1995) | | | | | | |
| | 3 (sites) | | | | | | |
| | Tsuchiya.E., Matsuzaki.G., Kurano.K., Fukuchi.T., Tsukao.A. and | | | | | | |
| | Miyakawa.T. | | | | | | |
| | The Saccharomyces cerevisiae SSD1 gene is involved in the tolerance | | | | | | |
| | to high concentration of Ca2+ with the participation of | | | | | | |
| | HST1/RRC1/BFRI | | | | | | |
| | Gene 176 (1-2), 35-38 (1996) | | | | | | |
| | 97075905 | | | | | | |
| | Location/Qualifiers | | | | | | |
| | 1. 4156 | | | | | | |
| | /organism="Saccharomyces cerevisiae" | | | | | | |
| | /strain="RAY-3a" | | | | | | |
| | /db_xref="taxon:4932" | | | | | | |
| | 946. 3693 | | | | | | |
| | /gene="HST1" | | | | | | |
| | 946. 3693 | | | | | | |
| | /gene="HST1" | | | | | | |
| | /standard_name="high-copy suppressor of TFP sensitivity" | | | | | | |
| | /codon_start=1 | | | | | | |
| | /protein_id="BA08819.1" | | | | | | |
| | /db_xref="GI:1816450" | | | | | | |
| | /translation="MSNRDNEMLRTTSSDKALASQDRKKESEVLLIAOSLIDNEIRSR | | | | | | |
| | KVLKRLTSGMDLIDPELDKEGSGSGRSGSPSKLNAVLKLLIWPANQHPNAPQ | | | | | | |
| | TPLELHGKNGSGIESNKTGQNGIYKRGVSPSKLNAVLKLLIWPANQHPNAPQ | | | | | | |
| | VAPDNFLVVDLTLLQSLSDNGEDNDGNSNENNDIEDNEKDESGSYENKENTIN | | | | | | |
| | NGSLSRHGNASLIRPSTLRSTFEDDNDNDKGDASAEYKVAERTSKITEREV | | | | | | |
| | SLDITTEELTKTNSAGLTNDNAITLARTLSMAGSYSDKRDQPPQESHIDBGDIGFST | | | | | | |

THIS PAGE BLANK (10/10)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 12, 2002, 10:38:12 ; Search time 4622.84 Seconds

(without alignments)
6512.739 Million cell updates/sec

Title: US-09-165-460A-1

Perfect score: 1825

Sequence: 1 acctaccctttttctatct.....aaaataacgattaacatt 1825

Scoring table: IDENTITY_NDC
Gapop 10.0 , Gapext 1.0

Searched: 1472140 segs, 8248589755 residues
Total number of hits satisfying chosen parameters: 2344280

Minimum DB seg length: 0
Maximum DB seg length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : GenEmbl:
1: gb_da:*
2: gb_hlg:*
3: gb_in:*
4: gb_cm:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_com:*
20: em_ov:*
21: em_ov:*
22: em_pat:*
23: em_ph:*
24: em_pl:*
25: em_ro:*
26: em_sts:*
27: em_sy:*
28: em_un:*
29: em_vl:*
30: em_htgo_hum:*
31: em_htgo_inv:*
32: em_htgo_rnd:*
33: em_htg_hum:*
34: em_htg_inv:*
35: em_htg_rnd:*
36: em_htg_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Match | Length | DB | ID | Description |
|------------|--------|-------|--------|----|-----------|----------------------|
| 1 | 1823.4 | 99.9 | 1825 | 8 | SCYR117W | 249617 S.cerevisiae |
| 2 | 1553.4 | 85.1 | 1706 | 8 | SCU7137 | 077137 Saccharomyces |
| 3 | 1342 | 18.7 | 1555 | 8 | SCYR116W | 249616 S.cerevisiae |
| 4 | 195 | 10.7 | 99360 | 8 | NCB11N2 | AL513444 Neurospor |
| 5 | 187 | 10.2 | 36920 | 8 | SPAC3H1 | 26814.4 S. pombe ch |
| 6 | 137.2 | 7.5 | 1275 | 8 | AF353722 | AF353722 Arabidops |
| 7 | 136 | 7.5 | 686 | 6 | A86270 | A86270 Sequence 92 |
| 8 | 135 | 7.5 | 686 | 6 | AR155763 | AR155763 Sequence |
| 9 | 135 | 7.5 | 686 | 6 | E66288 | E66288 genome DNA |
| 10 | 121 | 6.6 | 935 | 8 | SCYR118C | 249618 S.cerevisiae |
| 11 | 117.4 | 6.4 | 1859 | 9 | AF064867 | AF064867 Homo sapi |
| 12 | 117.4 | 6.4 | 2963 | 9 | AB016068 | AB016068 Homo sapi |
| 13 | 117.4 | 6.4 | 2966 | 9 | HSY13834 | Y13834 Homo sapien |
| 14 | 117.4 | 6.4 | 2966 | 6 | E32056 | E32056 Human ABC1. |
| 15 | 114.2 | 6.3 | 1340 | 3 | AF358443 | AF358443 Physarum |
| 16 | 98.4 | 5.4 | 78592 | 3 | AC004335 | AC004335 Drosophila |
| 17 | 98.4 | 5.4 | 102413 | 2 | AC020024 | AC020024 Drosophila |
| 18 | 92.8 | 5.1 | 41327 | 3 | CBRG47019 | AC084676 Caenorhab |
| 19 | 91.6 | 5.0 | 40767 | 3 | CEC04F12 | 281461 Caenorhabd |
| 20 | 98 | 4.8 | 317511 | 1 | CJ11168X3 | AL138076 Campyloba |
| 21 | 86.6 | 4.7 | 582 | 8 | CNS018P | AL110644 Botrytis |
| 22 | 81.6 | 4.5 | 660 | 9 | CNS018P | AL112927 Botrytis |
| 23 | 72.2 | 4.0 | 2161 | 9 | AK027874 | AK027874 Homo sapi |
| 24 | 66.8 | 3.7 | 1072 | 6 | AX055334 | AX055334 Sequence |
| 25 | 66.6 | 3.6 | 124808 | 2 | AP000652 | AP000652 Homo sapi |
| 26 | 65.6 | 3.6 | 16488 | 1 | AE004992 | AE004992 Halobacte |
| 27 | 65 | 3.6 | 256172 | 2 | AC005139 | AC005139 Plasmodiu |
| 28 | 64.4 | 3.5 | 169423 | 2 | AC069309 | AC069309 Mus muscu |
| 29 | 64.4 | 3.5 | 186306 | 2 | AC083946 | AC083946 Mus muscu |
| 30 | 63.6 | 3.5 | 1245 | 6 | AX055202 | AX055202 Sequence |
| 31 | 63.6 | 3.5 | 19715 | 1 | AE000555 | AE000555 Helicobac |
| 32 | 63.6 | 3.5 | 234112 | 3 | PFMAL1P2 | AL035475 Plasmodiu |
| 33 | 62.4 | 3.4 | 13383 | 3 | AE001374 | AE001374 Plasmodiu |
| 34 | 61.8 | 3.4 | 98734 | 2 | PFMAL1P2 | AL031745 Plasmodiu |
| 35 | 61.6 | 3.4 | 14943 | 1 | AE001528 | AE001528 Helicobac |
| 36 | 59.8 | 3.3 | 12029 | 3 | AE001392 | AE001392 Plasmodiu |
| 37 | 59.8 | 3.3 | 110000 | 2 | AC087899 | AC087899 Mus muscu |
| 38 | 59.2 | 3.2 | 114736 | 3 | PFMAL1P3 | 298547 Plasmodiu |
| 39 | 59.2 | 3.2 | 253305 | 3 | PFMAL1P7 | AL034559 Plasmodiu |
| 40 | 58.8 | 3.2 | 98899 | 2 | AL603743 | AL603743 Danio rer |
| 41 | 58.8 | 3.2 | 318221 | 2 | PFMAL13P3 | AL049184 Plasmodiu |
| 42 | 58.8 | 3.2 | 318503 | 2 | AC074166 | AC074166 Mus muscu |
| 43 | 58.4 | 3.2 | 172239 | 2 | AC069597 | AC069597 Homo sapi |
| 44 | 58.2 | 3.2 | 12029 | 3 | AE001412 | AE001412 Plasmodiu |
| 45 | 58.2 | 3.2 | 187469 | 9 | AL357312 | AL357312 Human DNA |

ALIGNMENTS

RESULT 1
SCYR117W
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
baker's yeast.
Saccharomyces cerevisiae
Eukaryote; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; Saccharomycetaceae; Saccharomyces.
SCYR117W 1825 bp DNA PLN 11-AUG-1997
249617 Y13136
249617.1 GI:1015836
REFERENCE
AUTHORS
JOURNAL
TITLE
JOURNAL
MIPS
Direct Submission
Submitted (25-SEP-1995) Data collected by MIPS on behalf of the
European yeast chromosome X sequencing project. MIPS at the
Max-Planck-Institut fuer Biochemie, Am Klopferspitz 18a D-82152

FEATURES
Location/Qualifiers
1. 1825

source

/organism="Saccharomyces cerevisiae"
/db_xref="taxon:4932"
/chromosome="X"

gene

CDS

/gene="SRE24"
343..1704
/gene="SRE24"
/note="ORF YJR117w"
/codon_start=1
/protein_id="CA89647.1"
/db_xref="GI:1015837"
/translation="MDPKTIIDHPNIPKIIISGSIASFESLYTROYKLSST
KLPVLEDEIDDETFHRSKNSRKAKREIFGDVYLAKLFIKIDLPKIHNAVS
LNNALVPRFHWSTVAOSLCEGLSSSTLDVLSTSHVLEKGFNLVQL
WTDNIRSLTAVIAGPILYLFKIFDEPFDLWYIVFVVOILMTIIPVIRI
PMENKTFPLEDELEKRSIESLADRVGFPDKIYFIDGSRSSSNAYFGLPFSKRI
VLEDTLVNSSTDEIPLAVLAHEIGHOKNHYMVIFESOLHTFLISLSTYSIRNSF
VYTPGFELKRSKSPVDPYITKEPFIILIGPMLFNIDILPELCMOPVMSLISRTHEYO
ADYAKKILGKONLCRALIDLOIKNLSTANVDPVLSYHISPTLAERLTALDYSEK
KKN"

BASE COUNT 570 a 397 c 280 g 578 t
ORIGIN

Query Match 99.9%; Score 1823.4; DB 8; Length 1825;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1824; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 acctaccctttttcttcttcaaacgaagccttcaacacacacacataactct 60
DB 1 acctaccctttttcttcttcaaacgaagccttcaacacacacacataactct 60
QY 61 acatatacatatacaatatacatatataacttgatattatattccttaacaaa 120
DB 61 acatatacatatacaatatacatatataacttgatattatattccttaacaaa 120
QY 121 aagaagcaatataacttctcttcttcttcttcttcttcttcttcttcttctt 180
DB 121 aagaagcaatataacttctcttcttcttcttcttcttcttcttcttcttctt 180
QY 181 tctgtctgtctgtcgcatttttccgagaaaaaatcgaggaataaaaaaagac 240
DB 181 tctgtctgtctgtcgcatttttccgagaaaaaatcgaggaataaaaaaagac 240
QY 241 aacgaacagagaaaaagcttcgaattataaacacattctataataacagagaaga 300
DB 241 aacgaacagagaaaaagcttcgaattataaacacattctataataacagagaaga 300
QY 301 aggaaaaaaagagaagaataagaaactcagagccttatcatgattgattcctaagag 360
DB 301 aggaaaaaaagagaagaataagaaactcagagccttatcatgattgattcctaagag 360
QY 361 attctcgacacataatacccggtggaattatcatattcttggttctcgattggccaa 420
DB 361 attctcgacacataatacccggtggaattatcatattcttggttctcgattggccaa 420
QY 421 ttttcttgaattcttcttcttcttcttcttcttcttcttcttcttcttcttctt 480
DB 421 ttttcttgaattcttcttcttcttcttcttcttcttcttcttcttcttcttctt 480
QY 481 ccaacctgtctgagaagcgaattgataatgaattcttcaataacgaagaatactacc 540
DB 481 ccaacctgtctgagaagcgaattgataatgaattcttcaataacgaagaatactacc 540
QY 541 cgggccaaggaagcgaattctcattctcgtagcttataacacgaagcccaaaagtaatt 600
DB 541 cgggccaaggaagcgaattctcattctcgtagcttataacacgaagcccaaaagtaatt 600
QY 601 ttcatacaataagcctcttccctaaatactcgacacatgagcgtttcttattgaaatga 660

DB 601 ttcatacaataatagcctcttccctaaatactcgacacatgagcgtttcttattgaaatga 660
QY 661 gtcctgacagtaagattatataatgctcctacgttcgcacagagttatagcttcggt 720
DB 661 gtcctgacagtaagattatataatgctcctacgttcgcacagagttatagcttcggt 720
QY 721 ccttatacagcttctcctacgttcggttcttcttcttcttcttcttcttcttctt 780
DB 721 ccttatacagcttctcctacgttcggttcttcttcttcttcttcttcttcttctt 780
QY 781 ctggaagaaaaatttgatttcaataaattgacccctcaacatagatgataatgatac 840
DB 781 ctggaagaaaaatttgatttcaataaattgacccctcaacatagatgataatgatac 840
QY 841 aagaagcttacttcttgagcgtatgcttcttcttcttcttcttcttcttcttcttctt 900
DB 841 aagaagcttacttcttgagcgtatgcttcttcttcttcttcttcttcttcttcttctt 900
QY 901 ttgataaattccctcctacgttcttcttcttcttcttcttcttcttcttcttcttctt 960
DB 901 ttgataaattccctcctacgttcttcttcttcttcttcttcttcttcttcttcttctt 960
QY 961 attctagcaatgaacatcatcctcagcttcaatcagcagtgatttaataagttcaactca 1020
DB 961 attctagcaatgaacatcatcctcagcttcaatcagcagtgatttaataagttcaactca 1020
QY 1021 ttgagagagggggaactgaaaaaatctattgaaatttgccgagatgagttggttccct 1080
DB 1021 ttgagagagggggaactgaaaaaatctattgaaatttgccgagatgagttggttccct 1080
QY 1081 ctgataaagatttttgatcttcttcttcttcttcttcttcttcttcttcttcttcttctt 1140
DB 1081 ctgataaagatttttgatcttcttcttcttcttcttcttcttcttcttcttcttcttctt 1140
QY 1141 acaagttcgcattcaccctcccaagagaattgcttcttcttcttcttcttcttcttctt 1200
DB 1141 acaagttcgcattcaccctcccaagagaattgcttcttcttcttcttcttcttcttctt 1200
QY 1201 tctactgatagaattcagcgttcttcttcttcttcttcttcttcttcttcttcttcttctt 1260
DB 1201 tctactgatagaattcagcgttcttcttcttcttcttcttcttcttcttcttcttcttctt 1260
QY 1261 atcgttaataatgctatccttcttcttcttcttcttcttcttcttcttcttcttcttctt 1320
DB 1261 atcgttaataatgctatccttcttcttcttcttcttcttcttcttcttcttcttcttctt 1320
QY 1321 agcatctacagaaatacatcatcttcttcaaacaccttcggttcttcttcttcttcttctt 1380
DB 1321 agcatctacagaaatacatcatcttcttcaaacaccttcggttcttcttcttcttcttctt 1380
QY 1381 ggcagttctgtatcccgcttcaactaagaaatcccaatcatcatgattgatttcttctt 1440
DB 1381 ggcagttctgtatcccgcttcaactaagaaatcccaatcatcatgattgatttcttctt 1440
QY 1441 tttaagacttataactcaccatcgaaatgagcagatgaaattctgagagatttattctcc 1500
DB 1441 tttaagacttataactcaccatcgaaatgagcagatgaaattctgagagatttattctcc 1500
QY 1501 agaacctcagaaatacaagcttgaattgcttcttcaaaaaattgagcctcacaacaaatcca 1560
DB 1501 agaacctcagaaatacaagcttgaattgcttcttcaaaaaattgagcctcacaacaaatcca 1560
QY 1561 tctagagctctaaattgatacacaatacaaaaacatttccacatgaaatgtagatcctctg 1620
DB 1561 tctagagctctaaattgatacacaatacaaaaacatttccacatgaaatgtagatcctctg 1620
QY 1621 tattctagatataatttccatccaaactcagctgtaaaagatcgacgctctgactat 1680
DB 1621 tattctagatataatttccatccaaactcagctgtaaaagatcgacgctctgactat 1680
QY 1681 gtttagtgaagaagaagaatactatctatagagtaacatatattagcattgacgtttaa 1740


```

FEATURES
source
1..1706
location/Qualifiers
/db_xref="taxon:9932"
/organism="Saccharomyces cerevisiae"
/chromosome="X"
/map="XJRI17W"
73..1434
/gene="STE24"
73..1434
/gene="STE24"
/function="first step of N-terminal processing of the yeast a-factor precursor"
/note="Ste24p"
/codon_start=1
/product="zinc metallo-protease"
/protein_id="AAB38271.1"
/db_xref="GI:1679741"
/translation="MFDTKTLIDHPNIIPWKLIIGSFIAQSFESYLTTRYOKLST
KLPLVLEIDDEFTHFSRNSRAKAFSITGDVYNLAOKLVEFKYDLFPKIMMAVS
LNNVLVPRFHMSTVAOSICFLGISLSTISLDVYPLSYSHFVLEEFKFNKTLVOL
MIPMISLTLAVALGIGLILFKIKDKRPTDPLMYMFLFVQVILAMITVETIM
PMFKPTPLEDGLKKSIESLADRYGPPLDKITVYDGSKSSHSNNAITFGIPTSKRI
VLEPTLVNSSTDEITAVLAHEIGHWOKNHLVNNVIFSQLTFLIFSLFTSYRNTSEF
YNTGFFLEKSTGSFVDPVITKEPPIILIGFMLFNDLLTFLPCAMQFVMSLISRHEVO
ADAVAKRLGYKONCRALIDIQINKLSIMMNDPLYSSYHSHPILARLRLADVDSEK
KKN"
BASE COUNT
494 a 366 c 275 g 571 t
ORIGIN

```

| | | | |
|----|------|---|------|
| QY | 331 | agacattatcatgcttgaatcttaagaacgattctcgcacatccctaattcccgigaa | 390 |
| Db | 61 | AGGCTTTATTCATGTTTATGATTTAAGACGATTCTCGACCAATCTTAATATCCGTGGAA | 120 |
| QY | 391 | ttaatcatctctcgtgctcgaatgcccaatttcttcgaactcttaactgaagctaa | 450 |
| Db | 121 | TTAATATTTCTGCGGTTCTCGATTGCCAATTTCTTTCGAATCTTACTTACGTACAGA | 180 |
| QY | 451 | cagtaacagaagctatcgtaaacaagtgtccacctgtgctgtgaagaagaaatgat | 510 |
| Db | 181 | CAGTACAGAGACTACTGAAACAAAGTGGCCACTGTGCTGAAGACGAATATGATAT | 240 |
| QY | 511 | gaacatttcataaatcaaggaactactctccggcgcaagggccaagtctccattt | 570 |
| Db | 241 | GAAACTTTTATATAATCAAGGAATACTCCCGGGCAAGGGCAAGTTCATTTTGGT | 300 |
| QY | 571 | gacgtcataactagcccaaaagctagtttttcaataacgaaccttcccaaatc | 630 |
| Db | 301 | GAGCTCATACTACCCAAAAGCTAGTTTTCATCAATACGACCTTCCCTAAATC | 360 |
| QY | 631 | tggcacatggcgttctcttaatgaatgcagtcctgcgaagcaagattcatatgtctc | 690 |
| Db | 361 | TGGCACAATGGCGCTTTCTTATTAATGAATGACAGTCGCGCAATTCAGATATGCTCC | 420 |
| QY | 691 | actgtgcagaaggttaatgctctctggctctcttaaccagtttgtctactgttgat | 750 |
| Db | 421 | ACTGTCCACAGAGTTAATGCTTCTTGGGCTCTTATTCAGTTGTCTACCTTGTTAT | 480 |
| QY | 751 | ttagcaactcttaactatagacatttgcctcggagaagaaattggtttcaataatg | 810 |
| Db | 481 | TTGGCACTCTTCTACTATACCATTTTGTGCTCGGAAGAAATTTGGTTTAATAATTG | 540 |
| QY | 811 | accgtccaactatgatacagcatatgaacaagaactgcacttgcgtacatgctatg | 870 |
| Db | 541 | ACGTCACAATAGATGATCAACGATATGATCAAGATCTGACTTGGGTATGATATGAT | 600 |
| QY | 871 | ggcccaactcttaaccgttcccttaagatcttgaataatccctactgatcttcc | 930 |
| Db | 601 | GGCCCAATCTTACTACCTGTTCTTAAAGATCTTGATTAATCCCTACATGATTTCTTGG | 660 |
| QY | 931 | tacatctgctctctctgttcgttgcgttcgaacttaagcatgacatcatctcagcttc | 990 |
| Db | 661 | TACATTTAGTCTTCTTGTGTCGTGTCACAACTTAGCATGACATATATCCAGTCTC | 720 |
| QY | 991 | atcatgcacatggttlaataagttcatctcatctatggagaacggtgaactgaataatc | 1050 |
| Db | 721 | ATCATGCGCATGTTTATATAGTCACTCATCTTGAGAGACGGTGAACGTAAATAATCTATT | 780 |
| QY | 1051 | gaaagtttgcgcgatagagttgggtctccctatagaatcttgcattgacgcgtca | 1110 |
| Db | 781 | GAAAGTTTGGCCGATATGATGGGTCCTCTAGATTAAGATTTTGTCTATGACGGCTCA | 840 |
| QY | 1111 | aaaagatcttctatccaagcagatcttcaagaagtttgcattacactccaagaagt | 1170 |
| Db | 841 | AAAAGATCTTCTCATTTCAATCAAGCATATTTCAAGGTTGCCATTACTCCAAAGATTT | 900 |
| QY | 1171 | gttttgcgcacacttaagtaaacagtaattctactga tgaatgaagcgtcttggcc | 1230 |
| Db | 901 | GTTTGTTCGACACTTATAGTAACAGTATTTCTACTGTAATATAGCGCTTTTGGCC | 960 |
| QY | 1231 | catgaatcgtgtaactcgtgcacaaaacaacacatcgtaataatgtaactcttaagcaatg | 1290 |
| Db | 961 | CATGAATCGGTACATGGCCAAAATAACCAATCGTTATATGTCATCTTTATGTCATATG | 1020 |
| QY | 1291 | caaacctctctaatttccaccttaccacagcatctcagaaataagaacttttaaac | 1350 |
| Db | 1021 | CACACCTTCTCATTTTCTCCCTTTACACACGATCTACAGAAATACATCTTTTACAC | 1080 |
| QY | 1351 | accttgcgttttcttaagaagaactgcactcgtcagtttgtgtatcccgttatacaatg | 1410 |
| Db | 1081 | ACCTTGGGCTTTTCTTAGAAGATGCACTGGCAGTTTGTGTGATCCGTTATACATAG | 1140 |
| QY | 1411 | gaattcccaatcatcatgattatgtaatttaacgactatttaactccaaccgaaatg | 1470 |

THIS PAGE BLANK (1.3PT0)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 12, 2002, 13:20:11 ; Search time 4622.84 Seconds

(without alignments)
10520.305 Million cell updates/sec

Title: US-09-165-460A-3

Perfect score: 2948
Sequence: 1 tgaactcttcatgaacaag.....ggggagagataagaatcaca 2948

Scoring table: IDENTITY-NUC
Gapop 10.0 , Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 08
Maximum Match 1008
Listing first 45 summaries

Database :

GenEmbl:*
1: gb.ba:*
2: gb.hg:*
3: gb.in:*
4: gb.om:*
5: gb.ov:*
6: gb.pat:*
7: gb.ph:*
8: gb.pl:*
9: gb.pr:*
10: gb.ro:*
11: gb.sts:*
12: gb.sy:*
13: gb.un:*
14: gb.vl:*
15: em.ba:*
16: em.fun:*
17: em.hum:*
18: em.in:*
19: em.om:*
20: em.or:*
21: em.ov:*
22: em.pat:*
23: em.ph:*
24: em.pl:*
25: em.ro:*
26: em.sts:*
27: em.sy:*
28: em.un:*
29: em.vl:*
30: em.hggo.hum:*
31: em.hggo.inv:*
32: em.hggo.rod:*
33: em.hggo.hum:*
34: em.hggo.inv:*
35: em.hggo.rod:*
36: em.hggo.other:*

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB | ID | Description |
|------------|--------|-------------|--------|----|-------------|--------------------|
| C 1 | 2948 | 100.0 | 29987 | 8 | SC8156 | 249260 S.cerevisia |
| 2 | 1560.4 | 52.9 | 4156 | 8 | D50276 | 249260 Saccharomy |
| 3 | 1361 | 46.2 | 4130 | 8 | YSCNR1G | 142821 Saccharomyc |
| 4 | 1163 | 39.5 | 4225 | 8 | SCU65682 | U65682 Saccharomyc |
| 5 | 1069.4 | 36.3 | 3250 | 8 | SCU32580 | U63849 Saccharomyc |
| 6 | 931 | 31.6 | 3525 | 8 | SCU63849 | D50849 Saccharomyc |
| 7 | 786.8 | 26.7 | 3240 | 8 | YSCBU1 | X88901 S.cerevisia |
| 8 | 703.8 | 23.9 | 3020 | 8 | SCRD1G8N | X88901 S.cerevisia |
| 9 | 225.2 | 7.6 | 270 | 8 | YSCGNC8X | 249210 S.cerevisia |
| 10 | 143.8 | 4.9 | 27559 | 8 | SC8339 | 249210 S.cerevisia |
| C 11 | 75 | 2.5 | 2023 | 8 | SCYNR06C | U71683 S.cerevisia |
| C 12 | 71.8 | 2.4 | 4775 | 8 | SCU32938 | U71683 S.cerevisia |
| C 13 | 68 | 2.3 | 780 | 11 | CNS06K8B | U71683 S.cerevisia |
| C 14 | 65.6 | 2.2 | 256172 | 2 | AC005139 | U71683 S.cerevisia |
| 15 | 65.6 | 2.2 | 310779 | 2 | AC005140 | U71683 S.cerevisia |
| 16 | 64.4 | 2.2 | 904 | 11 | CNS06189 | U71683 S.cerevisia |
| 17 | 64.4 | 2.2 | 86827 | 3 | PFMAL3P5 | U71683 S.cerevisia |
| 18 | 63.4 | 2.2 | 1920 | 3 | PFMAL3P5 | U71683 S.cerevisia |
| 19 | 63.4 | 2.2 | 110000 | 2 | PFMAL13P2_0 | U71683 S.cerevisia |
| 20 | 62 | 2.1 | 253305 | 3 | PFMAL13P7 | U71683 S.cerevisia |
| C 21 | 61.6 | 2.1 | 1141 | 6 | AX083744 | U71683 S.cerevisia |
| 22 | 60.8 | 2.1 | 224112 | 3 | PFMAL4P2 | U71683 S.cerevisia |
| 23 | 60.6 | 2.1 | 202872 | 2 | AC016160 | U71683 S.cerevisia |
| C 24 | 60 | 2.0 | 5632 | 3 | AF310892 | U71683 S.cerevisia |
| C 25 | 59.4 | 2.0 | 3214 | 3 | AF019980 | U71683 S.cerevisia |
| C 26 | 59.4 | 2.0 | 318221 | 2 | PFMAL13P3 | U71683 S.cerevisia |
| 27 | 58.8 | 2.0 | 321003 | 2 | PFMAL4P3 | U71683 S.cerevisia |
| 28 | 58.4 | 2.0 | 78439 | 17 | HS085197 | U71683 S.cerevisia |
| C 29 | 58.4 | 2.0 | 169546 | 2 | AC004157 | U71683 S.cerevisia |
| 30 | 58.4 | 2.0 | 245802 | 2 | AC006279 | U71683 S.cerevisia |
| 31 | 58.4 | 2.0 | 251124 | 2 | HD000660 | U71683 S.cerevisia |
| C 32 | 58.2 | 2.0 | 153098 | 3 | PFMAL3P2 | U71683 S.cerevisia |
| 33 | 58 | 2.0 | 159794 | 2 | AC004688 | U71683 S.cerevisia |
| 34 | 58 | 2.0 | 178157 | 9 | AC011302 | U71683 S.cerevisia |
| C 35 | 57.8 | 2.0 | 13684 | 3 | AE001403 | U71683 S.cerevisia |
| 36 | 57.8 | 2.0 | 153477 | 2 | AC006278 | U71683 S.cerevisia |
| 37 | 57.6 | 2.0 | 53932 | 2 | AC023371 | U71683 S.cerevisia |
| C 38 | 57 | 1.9 | 12029 | 3 | AE001431 | U71683 S.cerevisia |
| 39 | 57 | 1.9 | 153443 | 2 | AC006280 | U71683 S.cerevisia |
| C 40 | 57 | 1.9 | 17873 | 2 | AC005308 | U71683 S.cerevisia |
| C 41 | 57 | 1.9 | 196149 | 2 | AC004709 | U71683 S.cerevisia |
| C 42 | 56.8 | 1.9 | 67970 | 3 | PFMAL1P3 | U71683 S.cerevisia |
| C 43 | 56.8 | 1.9 | 149627 | 9 | AC087428 | U71683 S.cerevisia |
| C 44 | 56.6 | 1.9 | 92633 | 2 | PFMAL4P1_3 | U71683 S.cerevisia |
| C 45 | 56.6 | 1.9 | 153477 | 2 | AC006278 | U71683 S.cerevisia |

ALIGNMENTS

| RESULT | LOCUS | Accession | Definition | Version | Keywords | Source | Organism | Reference | Authors | Journal | Reference | Authors | Title |
|--------|----------|-----------|--|----------|-----------|--|--------------------------|----------------------|-----------------------------------|-------------|----------------------|-----------------------------------|-------------------|
| 1 | SC8156/c | 29987 bp | S.cerevisiae chromosome XIII cosmid 8156 | 249260.1 | GI:809081 | BUT1, cytochrome b5, DAG1, initiation factor 1A, inorganic pyrophosphatase, IPR2, nitrate reductase, Orotate phosphoribosyltransferase, tau element, TIF1, PRS3, transfer RNA-Gln, trehalose-phosphate synthase, U6 snRNP, URA10, baker's yeast. | Saccharomyces cerevisiae | 1 (bases 1 to 29987) | Barrell, B. and Rajandream, M. A. | Unpublished | 2 (bases 1 to 29987) | Barrell, B. and Rajandream, M. A. | Direct Submission |
| | SC8156 | 29987 bp | DNA | 249260.1 | GI:809081 | BUT1, cytochrome b5, DAG1, initiation factor 1A, inorganic pyrophosphatase, IPR2, nitrate reductase, Orotate phosphoribosyltransferase, tau element, TIF1, PRS3, transfer RNA-Gln, trehalose-phosphate synthase, U6 snRNP, URA10, baker's yeast. | Saccharomyces cerevisiae | 1 (bases 1 to 29987) | Barrell, B. and Rajandream, M. A. | Unpublished | 2 (bases 1 to 29987) | Barrell, B. and Rajandream, M. A. | Direct Submission |
| | SC8156 | 29987 bp | DNA | 249260.1 | GI:809081 | BUT1, cytochrome b5, DAG1, initiation factor 1A, inorganic pyrophosphatase, IPR2, nitrate reductase, Orotate phosphoribosyltransferase, tau element, TIF1, PRS3, transfer RNA-Gln, trehalose-phosphate synthase, U6 snRNP, URA10, baker's yeast. | Saccharomyces cerevisiae | 1 (bases 1 to 29987) | Barrell, B. and Rajandream, M. A. | Unpublished | 2 (bases 1 to 29987) | Barrell, B. and Rajandream, M. A. | Direct Submission |

COMMENT

Details of the omitted CDS are available on request. The more significant matches with motifs in the PROSITE database are also included but some of these may be fortuitous. The length of the motif, the number of codons in the motif, the codons and the calculated codon adaptation index (CAI) is given for each CDS.

FEATURES
SOURCE

```
complement(<1..356)
/note="YM8156.01c, unknown, len: 118, CAT: 0.14, overlaps
and extends YM9920.13c"
*
```

1. .140

```

/genes=11111
complement(851..,1312)
/genes="TTF11"

```

```

/product="Tlflp"
/protein_id="CA89243.1"

```

GRVEASCFCEDCNKRAHIRGKLRKKVWVGOGDIILVSLRDFODDQCDVHKYNLDEARF
LNKOGELPENAKINETDNFGESEDEVDNFEFGNADEDEDEEGEDELIDDI "

```

/gene="TPS3"
/note="YM8156.03c, TPS3 gene, len: 1054; CAI: 0.21,

```

```
sequence"
/codon_start=1
```

```

/db_xref="SWISS-PROT:P38426"
/translation="MTIVASLFLPYTPQEEADVTNSDTAKLVESSMIKVDCNNQEL"

```

TDAKTYYNPKPEGGYSSNAKLRLASLMSYSELKHLPTIVDSKGNGLSKNAVNIA
 KEELFQQQAFPSMAKVSJLAWDSYLAASSSLNRISSQFQNFIEDIUSDEUUSLSL
 AETVEEDVCSLGMCTPDELDEUENHCTCYRPODSCGRISSMDTMEGAVNVCT
 AETVEEDVCSLGMCTPDELDEUENHCTCYRPODSCGRISSMDTMEGAVNVCT

misc_feature

CDS

```

/note="YM8156.04, unknown", len: 313, CAI: 0.11, previously
noted ORF, SW: YTP3_YEAST P38430"
/codon_start-1

```

```

/ad_xref="SWISS-PROT:P38430"
/translation="MKKLVDÄHCHVITDPDNTFCGDDGSGQGLRCVMSNPNYDMMNL

```

TEOLOJYLAQAQARAWDEYIDASQAALRYENQOKFK" 6307 6302 6305

```

/product="unknown"
/protein_id="CA89246.1"

```

NNNNNNNSNNNGPTSSGRTNGKQQLTAAQQQYIKNLIETHITDNPDLPRSRBPM
DDEEYTDAPLRKYDHFQDLPVDNLTQGYLLGSKGAKTYSYKNTQGGHDKIKRR

SW:YTPL_YEAST_P38428, extends previously noted ORF".
/codon_start=1

/OD XIEL=SWISS-PROV:F38426
 /translation="MEDSRLLITLIVGVIFLKKKFSNQHPSPASORLSATGVNAHGR

2

```

completeness: 0.959,
/note="YM8156.07c, unknown, len: 461, CAT: 0.11"
/cool_start=1

```

```
/db_xref="SWISS-PROT:Q03508"  
/translation="MEFFFEFBPKGEWSEBPCGNHPTVPEKNSACETLVPEFVEEHA
```

LPRLKWSAIPVRPFTKSLPREQLELVNDMVRRKNSQALSTSQKEDLRRTPROIGIILL
CTASHRLLSIEVIKYLRKRGGRTRFRPKLYEHPILPCAEPGKIDIPALEIAKWSNND

```
CDS      9991. 12852.
/note="YM8156.08, unknown, len: 953, CAI: 0.20, probable
```

| | | | | | | |
|----|-------|-----------------|--|----------------------------|------------------|-------|
| QY | 481 | caacagagacattaa | ctgtgaacttggaa | tttaataaagacata | aggaaacttaattcc | 540 |
| | | | | | | |
| Db | 27305 | CAACGAGACATTAA | TGTGAACTTGGAATTTAA | TAAGACATTAAAGAACTTTAAATTC | | 27246 |
| QY | 541 | aagtttgaagcgtc | ctatgtttgaagttta | tgttcgttcagtaatga | tttaattga | 600 |
| | | | | | | |
| Db | 27245 | AAGTTTGAAGCTG | CCTATGTTGTAGCTTTAT | TGGCTTCGAAATATGATTTAAATTGA | | 27186 |
| QY | 601 | aaacacatcttggc | gttagcgaagattga | tatcccatcttcgtt | aggcaatgacaaaata | 660 |
| | | | | | | |
| Db | 27185 | AAACCACTTGCGG | TACGGAAGATGTGATATCCCATTTCTGTATGGCAAGACAATAA | | | 27126 |
| QY | 661 | aaaaaacttagaaa | aaatctctgttacc | tttcttaagatatagata | tgatggttc | 720 |
| | | | | | | |
| Db | 27125 | AAAAAATTAGAAAA | AAATTCCTGTATCTTTCTTAATAGATATAGATATATGTATGCTTT | | | 27066 |
| QY | 721 | gcttatagatgaag | tgattatctcgagcctt | gtgattcccatattata | aaatctctt | 780 |
| | | | | | | |
| Db | 27065 | GCTTATGATGAAG | ATATTATATGCGCGCTTGATATCCATATATTAATTAATTTCTTT | | | 27006 |
| QY | 781 | taaaatgcattctc | gtgtgctcttctgt | gtcttcgtatctt | tttttttttggacac | 840 |
| | | | | | | |
| Db | 27005 | TAAATGCATTTCT | GCTGCTCTTTTGTTGCTCTGATTTTTTTTTTTTTTGGACCACTG | | | 26946 |
| QY | 841 | gatggaaaacctt | tgatgattcttata | accttttttaacttcta | aaatcgaagttc | 900 |
| | | | | | | |
| Db | 26545 | GATGGAAACCTTGA | TGATTTATTAACCTTTATTTAAATTAACCTAAATAATCCAGATTTT | | | 26886 |
| QY | 901 | caggaaacaaaaca | tagatattctcttgc | aagaaaaaataaagaa | taaatgatgtctc | 960 |
| | | | | | | |
| Db | 26885 | CAGGAACAAACAT | AGATTTTCTTTGTCAAGAAAAATTAACGAATTAATTCATCTT | | | 26826 |
| QY | 961 | tgactactgacgtc | tcgtcatagagagaa | cagaaagcaatgc | tacaaattccaactt | 1020 |
| | | | | | | |
| Db | 26825 | TGACTACTGCTCT | GTCTCATAGAGAAACCAACACGCAATGCATCAATTCACAACTT | | | 26766 |

27 1021 CCGAGCGTCCCTAACACATCCCATCCGATGCTACCGCATATGCACCTTCAACC 26706
 D6 26765 TCTAGTGTCCCTAACATCTCCATTCCTAAGTGTGCTACCGCATATGCACCTTCAACC 26706

| | | | |
|----|-------|---|-------|
| QY | 1061 | agaagaggctcaaaacgaaataaacctccgaaacgatatcaatcccgatgaaaaaacttaaacat | 1140 |
| Db | 26705 | AGAAGGCTCAAAACGAAATATCATCTCGAACAGATTAATCTCGCATGCAAAAACCTTAACAT | 26646 |
| QY | 1141 | tatcgcatattccaaaccttttttggcgcttttttaacaatcccaatatcatctgtaccac | 1200 |
| Db | 26645 | TATGCTAATTCCAAACCTTTTGTGGCTTTTTCACATTCACATATCTGTACAC | 26586 |
| QY | 1201 | ttcacatatagtttcaaaagcagcatctttcaggcttagatcatcccaagttatacgc | 1260 |
| Db | 26585 | TTCACATATTAAGTTTCAAGAGCACCATTMTTAGCCTTAGGATATATCCAGGTTATTACGC | 26526 |
| QY | 1261 | tgcatctgccaacaccttgggaattcagccagctctgfgaaagaacttaagaatgctgttc | 1320 |
| Db | 26525 | TGCAATGCCAAACCCCTGGCAATTCAGCCAGTTCGGAAAGACTTAAACGAAATGCTGTGC | 26466 |
| QY | 1321 | gagcttatgcacctatactatctgtgacccggtttagaatttggatattatcatctattaaa | 1380 |
| Db | 26465 | GAGGTATTGACTTATATGTGTGACCCGTTTAGATTMTGATATATATATTATTTAAA | 26406 |
| QY | 1381 | ttccaaagagctatactctgaagaacttttaacaaatgcataatcttgaattcag | 1440 |
| Db | 26405 | TTCCAAAGAGCTCATCTCTGTAAGATTTTTCACATGATTCCTTAATATTTTGGAGTTTCAG | 26346 |
| QY | 1441 | gaatttatattctggacccaataacttagaagaatatcttaacgtcaatgctttgactac | 1500 |
| Db | 26345 | GAAATTTATATTGCCACCAATTACTGAGGAATATTTTAACGTCAATGCTTTTGAATAC | 26286 |
| QY | 1501 | gtactttaaacctaataccgcatctcgaacatacgaatcaacagttacttttggcaacatc | 1560 |
| Db | 26285 | GTACTTAAACCTAATACCGCATTCGCAACATACCTATCAACAGATTMTTTTGGCAACATC | 26226 |

| | | | |
|----|-------|--|-------|
| Db | 25145 | CCGGTAGAGAACTTCATGGAGGGGTAACTCAGGAGATAGAAATCCTCCAAATAGACTAA | 25086 |
| Qy | 2701 | caagtaactactcttaagtatataaaaaaggtgttcaactccatccaggaataatiaagct | 2760 |
| Db | 25085 | CAAGGTAACTACTTATGGATATATAAAAGGTGTCACTCTCCATCCAGGAATTTAAATGCT | 25026 |
| Qy | 2761 | aacgtatataaagaaaactcttctatctgtgtccgcacatcacacccctaagcttaagcct | 2820 |
| Db | 25025 | AACGTATTAAGAAATAACTTATTTATGTGGTCCCGCAATCAACCCCTTACGCTTAAGCCT | 24966 |
| Qy | 2821 | ataaatctcccaaggtctgttacaagataactcttaaaaaatatacactaagcgacaatgct | 2880 |
| Db | 24965 | GATATTTTCCTAGAGCTTGTAAGATACCTTTACAAATATATNCAACTAAGCACAATGCT | 24906 |
| Qy | 2881 | gaagataatactgttgggaatagacaatctgaaaaataacgatacttggagataatggggagataa | 2940 |
| Db | 24905 | GAAGTATATGATGGCAATAGCAATGAAAAATAACGATTTATGGAGATTAAGGGAGGATTA | 24848 |
| Qy | 2941 | gaatcacca 2948 | |
| Db | 24845 | CAATCACA 24838 | |

| | | | | | |
|-----------|---|--|-------------|-----|-------------|
| | RESULT | 2 | | | |
| D50276 | LOCUS | | | PJN | 10-FMB-1999 |
| D50276 | DEFINITION | D50276 | 4156 bp DNA | | |
| D50276 | ACCESSION | Saccharomyces cerevisiae HST1 gene, complete cds. | | | |
| D50276.1 | VERSION | HST1; high-copy suppressor of TFP sensitivity. | | | |
| | KEYWORDS | Saccharomyces cerevisiae (strain:RAY-3A) DNA. | | | |
| | SOURCE | Saccharomyces cerevisiae | | | |
| | ORGANISM | Eukaryote; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; | | | |
| | | Saccharomycetales; Saccharomycetaceae; Saccharomyces. | | | |
| | REFERENCE | 1 (bases 1 to 4156) | | | |
| AUTHORS | Tsuchiya,E. | | | | |
| TITLE | Direct submission | | | | |
| JOURNAL | Submitted, (13-APR-1995) to the DDBJ/EMBL/Genbank databases. Elko Tsuchiya, Hiroshima University Faculty of Engineering, Department of Fermentation Technology; Kagami-Yama, Higashi-Hiroshima, Hiroshima 739, Japan (E-mail:tetsuchi@ipc.hiroshima-u.ac.jp; Tel.:0824-24-7868, Fax:0824-24-7868). | | | | |
| REFERENCE | 2 (bases 1 to 4156) | | | | |
| AUTHORS | Tsuchiya,E., Matsuzaki,G., Tsukao,M. and Miyakawa,T. | | | | |
| TITLE | Molecular cloning and characterization of S. cerevisiae HST1 gene | | | | |
| JOURNAL | Unpublished (1995) | | | | |
| REFERENCE | 3 (sites) | | | | |
| AUTHORS | Tsuchiya,E., Matsuzaki,G., Kurano,K., Fukuchi.T., Tsukao,A. and Miyakawa,T. | | | | |
| TITLE | The saccharomyces cerevisiae SSDI gene is involved in the tolerance to high concentration of Ca ²⁺ with the participation of HST1/NMC1/BFRL | | | | |
| JOURNAL | Gene 176 (1-2), 35-38 (1996) | | | | |
| FEATURES | 97075905 | | | | |
| | Location/Qualifiers | | | | |
| | 1..4156 | | | | |
| | /organism="Saccharomyces cerevisiae" | | | | |
| | /strain="RAY-3A" | | | | |
| | /db_xref="taxon:4932" | | | | |
| | 946..3693 | | | | |
| gene | /gene="HST1" | | | | |
| CDS | 946..3693 | | | | |
| | /gene="HST1" | | | | |
| | /standard_name="high-copy suppressor of TFP sensitivity" | | | | |
| | /codon_start=1 | | | | |
| | /protein_id="BA08819.1" | | | | |
| | /db_xref="gi:1816450" | | | | |
| | /translation="MSNRDNESEMRLRTSSDKRAISQDRKRSEVLLIAQSLSINEIRSVL NKLRSLGSMDLIDPELDIKFGESGRSGAGTSASSAMSPDYTVNTYSDPP TKLETRRGNSGTGGESSKOGNYLGKKGVHSPRKLANYKNILAWPAQHNPENP VKPDFNELVDVTIANLOLSNGNDONGNSBNENDINDNCGEKSOSYEKENETINI NKGSRHNASLIIRPSFLRKSIIEFDENNEDKDGSASETYAKVERISKITERPY SLRDTEELRIINSAGLTNDATITMASYSODRKQPDPPEGHYDGDIIGFSST | | | | |

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 12, 2002, 13:20:11 ; Search time 4622.84 Seconds

(Without alignments)
10520.305 Million cell updates/sec

Title: US-09-165-460a-3

Sequence: 1 tgaactgtgtatgaacaaga.....ggggaggaataaatacaca 2948

Scoring table: IDENTITY_NTC

Gapop 10.0 , Gapext 1.0

Searched: 1472140 segs, 8248589755 residues

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seg length: 0

Maximum DB seg length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database:

GenBdb1:*
1: gb_da:*
2: gb_hgt:*
3: gb_in:*
4: gb_cm:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_da:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_om:*
20: em_or:*
21: em_ov:*
22: em_pat:*
23: em_ph:*
24: em_pl:*
25: em_ro:*
26: em_sts:*
27: em_sy:*
28: em_un:*
29: em_vl:*
30: em_hgt_hum:*
31: em_hgt_inv:*
32: em_hgt_rod:*
33: em_hgt_hum:*
34: em_hgt_inv:*
35: em_hgt_rod:*
36: em_hgt_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB | ID | Description |
|------------|--------|-------------|--------|----|------------|----------------------|
| C 1 | 2948 | 100.0 | 29987 | 8 | SC8156 | 249260 S.cerevisiae |
| 2 | 1560.4 | 52.9 | 4156 | 8 | D50276 | D50276 Saccharomyces |
| 3 | 1361 | 46.2 | 4130 | 8 | YSCNCRIG | Y49260 S.cerevisiae |
| 4 | 1163 | 39.5 | 4225 | 8 | SCU6582 | U65822 Saccharomyces |
| 5 | 1069.4 | 36.3 | 3250 | 8 | SCU32580 | U32580 Saccharomyces |
| 6 | 931 | 31.6 | 3525 | 8 | SCU6349 | U6349 Saccharomyces |
| 7 | 786.8 | 26.7 | 3240 | 8 | YSCBUL1 | D50083 Yeast BUL1 |
| 8 | 703.8 | 23.9 | 3020 | 8 | SCRDSTGEN | X88901 S.cerevisiae |
| 9 | 225.2 | 7.6 | 270 | 8 | YSCGWC58X | M87379 Yeast Eco R |
| 10 | 143.8 | 4.9 | 27559 | 8 | SC8338 | Y49210 S.cerevisiae |
| C 11 | 75 | 2.5 | 2023 | 8 | SCYNR06C | Z71683 S.cerevisiae |
| 12 | 71.8 | 2.4 | 4775 | 8 | SCU32938 | U32938 Saccharomyces |
| C 13 | 68 | 2.3 | 780 | 11 | CNS06KR8B | AL40261 T7 end of |
| 14 | 63.4 | 2.2 | 256172 | 2 | AC005139 | AC005139 plasmid |
| 15 | 65.6 | 2.2 | 310779 | 2 | AC005140 | AC005140 plasmid |
| 16 | 64.4 | 2.2 | 904 | 11 | CNS06T89 | AL39967 T3 end of |
| 17 | 64.4 | 2.2 | 86827 | 3 | PFMAL3P5 | AL034556 plasmid |
| 18 | 63.4 | 2.2 | 1920 | 3 | PFMAL3P7_0 | X62393 P.falciparum |
| 19 | 63.4 | 2.2 | 110000 | 2 | PFMAL3P2_0 | AL049185 plasmid |
| 20 | 62 | 2.1 | 253305 | 3 | PFMAL3P7 | AL034559 plasmid |
| C 21 | 61.6 | 2.1 | 1141 | 6 | AX083744 | AX083744 Sequence |
| 22 | 60.8 | 2.1 | 234112 | 3 | PFMAL4P2 | AL035475 plasmid |
| 23 | 60.6 | 2.1 | 202872 | 2 | AC016160 | AC016160 Homo sapi |
| C 24 | 60 | 2.0 | 5632 | 3 | AF310892 | AF310892 Dictyostea |
| C 25 | 59.4 | 2.0 | 3214 | 3 | AF019980 | AF019980 Dictyostea |
| 26 | 59.4 | 2.0 | 318221 | 2 | PFMAL3P3 | AL049184 plasmid |
| 27 | 58.8 | 2.0 | 321003 | 2 | PFMAL4P3 | AL035476 plasmid |
| 28 | 58.4 | 2.0 | 78439 | 17 | HS085197 | U85197 Homo sapien |
| C 29 | 58.4 | 2.0 | 169546 | 2 | AC004157 | AC004157 plasmid |
| 30 | 58.4 | 2.0 | 245802 | 2 | AC006279 | AC006279 plasmid |
| 31 | 58.4 | 2.0 | 251124 | 9 | HUAE006060 | AE006060 Homo sapi |
| C 32 | 58.2 | 2.0 | 153098 | 3 | PFMAL3P2 | AL034558 plasmid |
| 33 | 58 | 2.0 | 168794 | 2 | AC004688 | AC004688 plasmid |
| 34 | 57.8 | 2.0 | 178137 | 9 | AC011302 | AC011302 Homo sapi |
| C 35 | 57.8 | 2.0 | 13684 | 3 | AE001403 | AE001403 plasmid |
| 36 | 57.8 | 2.0 | 153477 | 2 | AC006278 | AC006278 plasmid |
| C 37 | 57.6 | 2.0 | 53932 | 2 | AC023371 | AC023371 Homo sapi |
| C 38 | 57 | 1.9 | 13029 | 3 | AE001431 | AE001431 plasmid |
| 39 | 57 | 1.9 | 163443 | 2 | AC006280 | AC006280 plasmid |
| C 40 | 57 | 1.9 | 178273 | 2 | AC005308 | AC005308 plasmid |
| C 41 | 57 | 1.9 | 196149 | 2 | AC004709 | AC004709 plasmid |
| C 42 | 56.8 | 1.9 | 67970 | 3 | PFMAL3P3 | AL031746 plasmid |
| C 43 | 56.8 | 1.9 | 149627 | 9 | AC087428 | AC087428 Homo sapi |
| C 44 | 56.6 | 1.9 | 92633 | 2 | PFMAL4P1_3 | Continuation (4 of |
| C 45 | 56.6 | 1.9 | 153477 | 2 | AC006278 | AC006278 plasmid |

ALIGNMENTS

RESULT 1
SC8156/c
LOCUS SC8156 29987 bp DNA PLN 11-AUG-1997
DEFINITION S.cerevisiae chromosome XIII cosmid 8156.
ACCESSION Z49260 Z71257
VERSION Z49260.1 GI:809081
KEYWORDS BUL1; cytochrome b5; DNGL; initiation factor 1A; inorganic pyrophosphatase; IPP2; nitrate reductase; orotate phosphoribosyltransferase; tau element; TIF1; TP3; transfer RNA-Glu; trehalose-phosphate synthase; U6 snRNP; URA10.
SOURCE Baker's yeast.
ORGANISM Saccharomyces cerevisiae
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; Saccharomycetales; Saccharomycetaceae; Saccharomyces.
REFERENCE 1 (bases 1 to 29987)
AUTHORS Lye, G. and Churcher, C.M.
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 29987)
AUTHORS Barrett, B. and Rajandream, M.A.
TITLE Direct Submission

THIS PAGE BLANK (CPTO)

JOURNAL

COMMENT

Submitted (12-MAY-1995) *Saccharomyces cerevisiae* chromosome XIII sequencing project, Sanger Centre, Hinxton Hall, Hinxton, Cambridge CB10 1RQ. E-mail: barrell@sanger.ac.uk

NOTES:

All CDS over 100 codons have been analysed. CDS that are completely overlapped and those that are overlapped by more than 50% of their length by a larger CDS have been omitted from this analysis. Details of the omitted CDS are available on request. The more significant matches with motifs in the PROSITE database are also included but some of these may be fortuitous. The length in codons and the calculated codon adaptation index (CAI) is given for each CDS.

Cosmid 8156 is overlapped at the start by cosmid 9920, emb1 entry SC9920, accession no. Z48639 and at the end by cosmid 8021.

FEATURES

source

Location/Qualifiers
1. 29987

/organism="Saccharomyces cerevisiae"

/strain="AB972"

/db_xref="taxon:4932"

/chromosome="XIII"

/clone="cosmid 8156"

/map="13R"

/complement(<1. 356)

/note="YM8156.01c, unknown, len: 118, CAI: 0.14, overlaps and extends YW920.13c"

/codon_start=1

/product="unknown"

/protein_id="CA89242.1"

/db_xref="GI:809082"

/db_xref="SWISS-PROT:Q03496"

/translation="MTDVQSSQSEIELFRYKEFLIANNPKINNENKDAVLIQIEHD ERYLIQIKDGLPMLNLESTRLIFPDTSCILRSHQIILASRKIDSOEFLSIVKQQLLT EANNIIFEEYLDLFWA"

1. 140

/note="overlap with SC9920 Z48639 S. cerevisiae chromosome XIII cosmid 9920"

/complement(851. 1312)

/gene="TIF11"

/complement(851. 1312)

/gene="TIF11"

/note="YM8156.02c TIF11 gene, IF1A_YEAST, P3891, initiation factor 1a"

/codon_start=1

/product="Tif1p"

/protein_id="CA89243.1"

/db_xref="GI:809083"

/db_xref="SWISS-PROT:P38912"

/translation="WGKRTGKGRGKGRGKNDSDGPKRELLYKEGGEYAOITMLGN GRVESCDDGNKRAHRIKGLAKTYMGQGLIIVSLDPODDCDVYKYNLEART LKNGELPENAKINETDNGFESDVAVFEGMDEDEDEDELIDDI"

/complement(1678. 4842)

/gene="TPS3"

/complement(1678. 4842)

/gene="TPS3"

/note="YM8156.03c, TPS3 gene, len: 1054, CAI: 0.21, SW:TPS3_YEAST P38425, alpha.alpha.-trehalose-phosphate synthase and PIR:S27471, P500213 Lipocalin signature; translation in TPS3_YEAST starts at amino acid 33 in this sequence"

/codon_start=1

/product="Tps3p"

/protein_id="CA89244.1"

/db_xref="GI:809084"

/db_xref="SWISS-PROT:P38426"

/translation="MTIIVASFLPYTPPEADVTNSDTAKLVSSMLKIVDCCNOELS NNKORSSSVTSASHYIGLPQEAQINGEPLRVANGSPATGVNHEMELSSQFL EELTANATHAASGIPRANPPYSSGSTRAPRVEVEFPAPARVCSPOQASASISA SRSASHNDLSSSLMKNPVLSFDSHPVRSSKSVITPVKSVDDPAVDVAKY REFOQQAFLPMKRVSGTAGDSSIASSSNMLRISQOFQDNFIDTSDSEDIDLE TDAIKRYNPKGKYSNNAKLRASLRNSYELFKHLPMTIVDSKNGSKNAIVIAV AKTYKEPVSWGTGIGTIDELPHEVCHRSIKKLEDDFSFPVTDITFKAYKNA

AKTYKEPVSWGTGIGTIDELPHEVCHRSIKKLEDDFSFPVTDITFKAYKNA

misc_feature

CDS

/note="YM8156.04, unknown, len: 313, CAI: 0.11, previously noted ORF, SW:YTP3_YEAST P38430"

/codon_start=1

/product="unknown"

/protein_id="CA89245.1"

/db_xref="GI:809085"

/db_xref="SWISS-PROT:P38430"

/translation="NKKLVDAHCHVITDPTDFCCGDDGSGQGLRCVMSNPYDMNTL KLAGSTSKNDICVGFVHPWVSHLYGSRDKVSHODVLEYKNEBOFDSLYOYL PPDLLEKIRKRPDLVSYVIGIGDLPFLPANGFMONKRAVLTAKYLSHOE TVRRFRWLARHRSKPTISLHDYCKHCKLNDICNEELLTHSTVCKHSTTGKPTLLG QWLKRFPPDRIFSLSKWLNFDPEGDLVSLPSTCLITETDYPIDNPDSYOKAL TEOLOYLNQOIRAWDETLDASQALRYENQKFLK"

6392. 6997

/note="YM8156.05, unknown, len: 301, CAI: 0.12, SW:YTP2_YEAST P38429 previously noted ORF"

/codon_start=1

/product="unknown"

/protein_id="CA89246.1"

/db_xref="GI:809086"

/db_xref="SWISS-PROT:P38429"

/translation="NARPVYNAETSSRGRPTGCGGYASNNNSCANNNSNNNNNN NNNNNNSNNNGPTSSGRNKGOLTAQOQYIKNLLETHITDHTDPSHSPHA DPEYTDAPLRKYDHFOLDVDPNLTGLGILGSKYAKYSTRKNOGQDHRIRER DLANVRRHRHDEHSIKETDCIPQFIYKVKKKKFKFEPFG"

7278. 7889

/note="YM8156.06, unknown, len: 301, CAI: 0.13, SW:YTP1_YEAST P38428, extends previously noted ORF"

/codon_start=1

/product="unknown"

/protein_id="CA89247.1"

/db_xref="GI:809087"

/db_xref="SWISS-PROT:P38428"

/translation="MDSRLTLITLVFVIFLKFQSNQHPAQRSLATGVNAHGR POGSTNALARTGVNGHVTOMVETVONLAPNLHPEDQIRLSLETGSTEIVERT LGDESEFPGEPSRAPMGANAVDNNAGGEGFPKMKMCAEYTLDFHVDLN EDMNSLTFDQDIEERRRLIWOARKKLEFKLOSQDQSLLT"

/complement(8013. 9398)

/note="YM8156.07c, unknown, len: 461, CAI: 0.11"

/codon_start=1

/product="unknown"

/protein_id="CA89248.1"

/db_xref="GI:809088"

/db_xref="SWISS-PROT:Q03508"

/translation="MEFEDEFRRKRELSGCGNRVLRKNSCACELAPEVEBHAI TKLIYDFDNTLFPATPGTQLATYRELNLITSLTPNGGWNPGFLOALIEISIKT PRRYSWADIVKLAEESYSAKDTISYLGRESEKFKLIEHALOTAFSHKCSENE FPNVAVCAKRAISSEYTSKYEKLMDLLEVPYSLRELSYDDRIHOIDAFSEFSSID LPRKNSAIVRPFPTKSLPREQDLEAMDMVRKNSCALSTSQFDLARTROICYIL CTASHRLSLEVIKYLKRRGRGTRFKRLIHPYLICAEAGKQIPALELAKVNSND TRPDSSEKQIHOIQLIYLEDPQKCIYHFPVDTVASAHNNRKPDELYFKVLEP NRRTFTEPEYITVGHYKRRDRIDLEVTERLINCXEDIHVWLDWTIPKAFGRF AKLAATICSNA"

9991. 12852

/note="YM8156.08, unknown, len: 953, CAI: 0.20, probable integral membrane glycoprotein"

/codon_start=1

/product="unknown"

/protein_id="CA89249.1"

/db_xref="GI:809089"

/db_xref="SWISS-PROT:Q03509"

/translation="MEFEDEFRRKRELSGCGNRVLRKNSCACELAPEVEBHAI TKLIYDFDNTLFPATPGTQLATYRELNLITSLTPNGGWNPGFLOALIEISIKT PRRYSWADIVKLAEESYSAKDTISYLGRESEKFKLIEHALOTAFSHKCSENE FPNVAVCAKRAISSEYTSKYEKLMDLLEVPYSLRELSYDDRIHOIDAFSEFSSID LPRKNSAIVRPFPTKSLPREQDLEAMDMVRKNSCALSTSQFDLARTROICYIL CTASHRLSLEVIKYLKRRGRGTRFKRLIHPYLICAEAGKQIPALELAKVNSND TRPDSSEKQIHOIQLIYLEDPQKCIYHFPVDTVASAHNNRKPDELYFKVLEP NRRTFTEPEYITVGHYKRRDRIDLEVTERLINCXEDIHVWLDWTIPKAFGRF AKLAATICSNA"

9991. 12852

/note="YM8156.08, unknown, len: 953, CAI: 0.20, probable integral membrane glycoprotein"

/codon_start=1

/product="unknown"

/protein_id="CA89249.1"

/db_xref="GI:809089"

/db_xref="SWISS-PROT:Q03509"

/translation="MEFEDEFRRKRELSGCGNRVLRKNSCACELAPEVEBHAI TKLIYDFDNTLFPATPGTQLATYRELNLITSLTPNGGWNPGFLOALIEISIKT PRRYSWADIVKLAEESYSAKDTISYLGRESEKFKLIEHALOTAFSHKCSENE FPNVAVCAKRAISSEYTSKYEKLMDLLEVPYSLRELSYDDRIHOIDAFSEFSSID LPRKNSAIVRPFPTKSLPREQDLEAMDMVRKNSCALSTSQFDLARTROICYIL CTASHRLSLEVIKYLKRRGRGTRFKRLIHPYLICAEAGKQIPALELAKVNSND TRPDSSEKQIHOIQLIYLEDPQKCIYHFPVDTVASAHNNRKPDELYFKVLEP NRRTFTEPEYITVGHYKRRDRIDLEVTERLINCXEDIHVWLDWTIPKAFGRF AKLAATICSNA"

9991. 12852

THIS PAGE BLANK (OPTION)

THIS PAGE BLANK (OPTION)

```

13245..14177
/gene="ipp2"
13245..14177
/gene="ipp2"
/note="YM8156_09, IPP2 gene (PPA2), len: 310, CAI: 0.14,
SM: pPY2_YEAST, P28239, inorganic pyrophosphatase,
incomplectid, contains P80037 Inorganic pyrophosphatase
signature"
/codon_start=1
/product="Ipp2p"
/protein_id="CA89250.1"
/db_xref="GI:809090"
/db_xref="SWISS-PROT:P28239"
/translation="MTLRNNATLSKARSRLEERKQTLNLISRNRRQSTIQGSKTT
GKKKTLILNGEYGFHVPDLNDEHETVMATVEVPMWTGFEISKELRNPTV

```

| | | | |
|----|-------|--|-------|
| OY | 1 | tgactgtttgaagaacaagaagaagcttgacaagcatcaaaagcttttaagatgattcca | 60 |
| Db | 27785 | tgaactgtttgaatgacacaaagagagagctgacaaacatcaaaagctttttagaagattcca | 27728 |
| OY | 61 | ttcaaaaatctgtgaatcatgatgaacaagaftcacaagaatttctctgaattlaaatgagt | 120 |
| Db | 27725 | ttcaaaaaatttttgatattgaaaccagaftccaagaattttcttgatataatgagtt | 27666 |
| OY | 121 | ataataatagaattaggagagaccgtgtggccaagaagaactgaattcaagatttttatc | 180 |
| Db | 27665 | atataatatagaattaggagagaccgtgtggccaagaagaactgaattcaagatttttatc | 27606 |
| OY | 181 | ttcacacgctgttttaacagatatcgaaagcaattctgcaacttgaaagttagtgcacaact | 240 |
| Db | 27605 | ttcacacgctgttttaacagatatcgaaagcaattttgcaacttgaaagttagtgcacaact | 27546 |
| OY | 241 | atcccaacattttttaaanaaacaggtctgatacccttaaaacaacactcaagaagcattatc | 300 |
| Db | 27545 | atcccaacattttttaaanaaacaggtctgatacccttaaaacaacactcaagaagcattatc | 27486 |
| OY | 301 | tgaagattcaataatcgcacacaggtaaagtagttatcatcgctgccagttccagtcatt | 360 |
| Db | 27485 | tgaagattcaataatcgcacacaggtaaagtagttatcatcgctgccagttccagtcatt | 27422 |
| OY | 361 | aacgcagcgaatactcttcatccaagaagtcagttatttttaactagcgtagctcgtctac | 420 |
| Db | 27425 | aacgcagcgaatactcttcatccaagaagtcagttatttttaactagcgtagctcgtctac | 27366 |
| OY | 421 | ttccctcgaaatttaacagaccagaatttcaataatggttagattgctccctttaagta | 480 |
| Db | 27365 | ttccctcgaaatttaacagaccagaatttctcatataatggattgagattgctccctttaagta | 27306 |

| | | | |
|----|-------|---|-------|
| QY | 481 | caacagacacatlaatggaacttggaaacttaataagaacatlaaggaacttaattcc | 540 |
| Db | 27305 | CAACGACACATTAATGTGAACCTGGAAATTAATAGACATTAAGGAACCTTTATATCC | 27248 |
| QY | 541 | aagtttgaagcgtccatgcttgaagtttcatgctgcgtgaagatgaatlaattgga | 600 |
| Db | 27245 | AAGTTTGAAGCGTCCATGTGAGTTTATTTGCGCTTGAGAAATGATTAATTTTGA | 27168 |
| QY | 601 | aaaccatcttggcgtacggaagatgatacccatattctgttaggcaagtacaaaata | 660 |
| Db | 27185 | AAACCATTCTGGCGAGGGAAGATGATATCCCATTTCTGTAGGCAAGTCACAAAATA | 27128 |
| QY | 661 | aaaaaaccttgaaaaaaattctgccttaactttcttatagatatagatatgtagt | 720 |
| Db | 27125 | AAAAAACCTTGAAAAAAATTTCTGTAATTTTCTTAATAGATATATATATATATGTTT | 27066 |
| QY | 721 | gcttaataatgaagatlaattacgcgtcccttgtatcccatataataaatctt | 780 |
| Db | 27065 | GCTTAATATATAGATATTTATTCGGTCTCTTGATATCCCATATTTATTAATAATCTTT | 27006 |
| QY | 781 | taaatgcatcttctgtgtctcttltgttctgtatcttcttcttcttcttcttctt | 840 |
| Db | 27005 | TAAATGCAATTTTCGCGCTCTTTGTGTCCTGATTTTCTTTTCTTTTGGACACATG | 26948 |
| QY | 841 | gatggaaaaccttgaatgatttattacttattttaaagttaacaaatatcgagatt | 900 |
| Db | 26945 | GATGGAACCACTTGTATATTTTATTAACCTTATTTAATGATACATAAATACGAGATT | 26886 |
| QY | 901 | caggacaacaaatagaattcttctgtfccaagaaaataaaacggaataaatgtgct | 960 |
| Db | 26885 | CAGGACAAACATATGAATTTTCTTTGTCAAGAAAAATTAACGAAATTAATTTGATCTT | 26822 |
| QY | 961 | tgaactacgtactgtctgcataagagaacagaacagaatgcatacttccaaatt | 1020 |
| Db | 26825 | TGACTACGACGTGTGCATAGAGAGAACAGACAGCAATGCTATTCATTTCTCAACAT | 26766 |
| QY | 1021 | tctagtctccatacatctccatactcctatgtgtctaccgctatactgaacttcaacc | 1080 |
| Db | 26765 | TCTAGTCTCCTCATATCACTCTCCATTCCTATGTGCTACCGGTAATGCAACTTCCAAAC | 26706 |
| QY | 1081 | agaagagcttcaagagaatactccctcggaagatataactctgcacagcaaaaacttaacat | 1140 |
| Db | 26705 | AGAAAGCTCTTAACGAGATATCTCGAAGCATTAATCTCGCATGCAAAAACTTACAT | 26648 |
| QY | 1141 | tatgtcaattcccaaccccttcttgtgtcccttcttcttcaaatcccaatccatagaccac | 1200 |
| Db | 26645 | TATGCTAATTTCCAAACCTTTTGTGTGCTTTTAAATCTCAATCTATCTAGTACAC | 26588 |
| QY | 1201 | tccacataaagttccaagagcaatcttcttaagcttaagatcatccaggttaatacgc | 1260 |
| Db | 26585 | TTCACATTAATGATTTCAAGAGCGCATTTTATGCGCTTAGATTAATCCACAGGTATTACGC | 26528 |
| QY | 1261 | tgcattgcacaaaccttggcaattcaagcagaattctgtgaagaacttaagaaatgtgtgc | 1320 |
| Db | 26525 | TGCATTGCCAAACCTTGGCAATTCAGCCAGTTCGTGAAGCTTAAGCAAAATGTTGTGC | 26466 |
| QY | 1321 | gatgttattgacctatacttvgaacccgttcttagattctgtatataatcaatttaaa | 1380 |
| Db | 26465 | GATGTATTGACCTATATGTGAGACCGGTTTAAATTTGTATATATATCAATTTATTA | 26406 |
| QY | 1381 | tccaaagagccctatacttgaagatttttccatgaatctctgaatcttggagttcaag | 1440 |
| Db | 26405 | TCCAAAGAGCCCTATATCTTGAAGATTTTACCATTAATCTCGAATATTGTGAGTTTACG | 26346 |
| QY | 1441 | gaatttatatttggacaataacgtaggaataattttaacgcgtcaagccttttaccac | 1500 |
| Db | 26345 | GAAATTTATTTTGCACCAATACTGAGGAATAATTTTACACGTAAAGCTTTTAACATAC | 26286 |
| QY | 1501 | gtacttaaaccttaaccgaatcttgcacactaagcatacaacagttatttggcaaccac | 1560 |
| Db | 26285 | GTACTTAACCTTAATACGCAATGCGCAATGCGCACTAAGCATCAACAGTTATTTGGCAACATC | 26226 |
| QY | 1561 | gcttttttcttggacttgcacgcacacacatgcttataggaatatacgaagaagctcat | 1620 |

THIS PAGE BLANK (0.5PT0)

THIS PAGE BLANK (0.5PT0)

Db 26225 GCTTTTGGACTTGGCAGCAGACACCATGCTTATGACCAATTCAGAAAGGCTCCAT 26166
 Qy 1621 gacaactgttccatctcgtcgcacaaatgtcccaattttacacaaacacttttgg 1680
 Db 26165 GACAACGTGTTCCATCTGCTGACACACATGCTTCCAAATTTTATACCAACATTTTGG 26106
 Qy 1681 agggtaacaaattgtatctgtaagaacagcgcggaacatgtgtgcgcataaccc 1740
 Db 26105 AGGGTTAACCAAGTTTGTATTCGTAACACAGCGGGAACCTAGTGGCTGACATTAACCT 26046
 Qy 1741 gactgccccttgcaatacgaatggtccctgctccttcaagatgtaattacattcac 1800
 Db 26045 GCATGCCCTTTCGAATTCAGGGGTTCCGGTCCCTTCAAGATGTAATTCATTTAC 25986
 Qy 1801 agtagtagacaagaagcgtgagcgaattcccaattgtgtcccaatcgtgaatggtcta 1860
 Db 25985 AGTAGTAGACAAGAAGGAGCAGCATTTCCAAATTTGCTCAATCTGGAATTAAGTGCTA 25926
 Qy 1861 ctctgcacgtcgtgctccttgatataatccctgaagaatccctacaaactcgtgtag 1920
 Db 25925 CTTCGCACTGCTGCTGCTGATTAATATCCCTGAAGATACCTTACAACTCTGCTAG 25866
 Qy 1921 aaccctgtgtatagaataaacccttagcctttttacgtactgtataccgtttaaat 1980
 Db 25865 AACTCTGCTGTATAGAAATACCCCTTAGCCTTTTACGACTGTATACCGTTAAAT 25806
 Qy 1981 ttccatgtactataacacttttttcaactataatgaaatgtcgcagagcggcgt 2040
 Db 25805 TTCCATGTAGCATTAACCTTTTTCACACTATATGAAATCTATCGACGACCGGCGCT 25746
 Qy 2041 ttgttgcaggaagtgaaataacgtggtttgtgtgtttgtgtgaagaatttggagac 2100
 Db 25745 TTGTGTACCGAAGAGTGAATAAATCAGATTTGTTGTTGTAAGAAATTTGGAGAC 25686
 Qy 2101 tetaagttaccataacttctgtatgaactacataacagtcgttgcgttagtggat 2160
 Db 25685 TATAAAGTACCTTACTTTGTTATACGACTCAATACAGTGGTGGTGTAGGTAT 25626
 Qy 2161 tgaagttgcagatctcaagaagtagagaaggtgtgcataatagtttcgcagcttttc 2220
 Db 25625 TGAAGTTGCAGATCTAAGAGTAGAAGAGGAGGATCAATTAAGTTTCGAGGTTTTC 25566
 Qy 2221 ttttttcaaggttttcaattgtgtctcctagaatttaaggtccttagtggttgt 2280
 Db 25565 TTTTAAAGGTTTATTTGTTGCTCTCAATTAAGGTTCTAGTTAGTTGGTTG 25506
 Qy 2281 ttgtgtgttaccatattttcaattcaagaaggaatttagtgcctttataatgtccat 2340
 Db 25505 TTGTGGGTTACATATTTTCAATTCAGAGGAATTAAGGCTTTTATAAGTCCAT 25446
 Qy 2341 agagatacagagcagtcgtcgtactacatcaagcgataagcgatcgctagtcaag 2400
 Db 25445 AGAGATAACGAGACATGCTGCTGACTACATCAAGCATTAAGGAGCGGTAGTCAAG 25386
 Qy 2401 gataaacggaagtcgaagtttgaattgtcgtcacagcgccttgcaatgaaatccgagc 2460
 Db 25385 GATTAACGGAAGTCTGAAGTTTGTATGCTGCAAGTCCCTTGAATTAAGAAATCCGAC 25326
 Qy 2461 gtaaaaaaacctaaaaaatgtcgattgtgtcgaatgtaattactattgtaccagaat 2520
 Db 25325 GTAAAAAACCTAAAAAGTTTGTGATTTGGTCAATGATTTACTTATGATCCGAATTA 25266
 Qy 2521 gataaaatcgttggggaatctagtgtggaagcagatcgtgtgcagacatccagt 2580
 Db 25265 GATTAATAATTCGTTGGGGAATCTAGTGGAGACATCATGCTGCGACACATCCAGT 25206
 Qy 2581 tctgcgtcaatgcgaatgtgacacacacccgttaataacacagatatagatccact 2640
 Db 25205 TCTGCGCTCAATGCGAAGTACACACACCGCTTATTAATACACAGATTAAGCATCACT 25146
 Qy 2641 ccgctagaagacttgcaggggggttaactcaggagataagatcctccataagaactaa 2700

Db 25145 CCGTAGAAGACTTGATGAGGAGGGGTAACCTAGGATTAAGTAATCTCCATTAAGCTAA 25086
 Qy 2701 caagtaactacttggatataaaaaagtttcaactcctccatcagaatgaatgtc 2760
 Db 25085 CAAGTAACCTACTTGGATATAAAAAAGGTTCCTCTCTCATCCAGGAATTAATAGCT 25026
 Qy 2761 aacgtatnaaagaactatattatgtgtccgcgaacaaacacacttaagtc 2820
 Db 25025 AACGTATTAAGAAAACCTTATTAATGAGTCCCGCAATCAACACCTTAAGTTAGCCT 24966
 Qy 2821 gataattcctagagctgttacaagatacttatacaaatatacaactaagcgaatgt 2880
 Db 24965 GATAATTTCTAGAGCTTTACAGATTAATTAACAAATATACAACTAAGCGCAATGCT 24906
 Qy 2881 gaagataatgagtggaatgcaatgcaataacgaatattgaggaatgaggaataa 2940
 Db 24905 GAAGATAATGANTGGAAATGACATGAATAATAGATTTGAGATTAAGGAGGATTA 24846
 Qy 2941 gaatacga 2948
 Db 24845 GAATCACA 24838

RESULT 2

LOCUS D50276 4156 bp DNA PLN 10-FEB-1999
 DEFINITION Saccharomyces cerevisiae Hst1 gene, complete cds.
 ACCESSION D50276
 VERSION D50276.1 GI:1777313
 KEYWORDS HST1, high-copy suppressor of TRP sensitivity.
 SOURCE Saccharomyces cerevisiae (strain: RAY-3a) DNA.
 ORGANISM Saccharomyces cerevisiae
 Eukaryote; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 Saccharomycetales; Saccharomycetaceae; Saccharomyces.

REFERENCE 1 (bases 1 to 4156)
 Tsuchiya, E.
 Direct Submission
 Submitted (13-APR-1995) to the DDBJ/EMBL/GenBank databases. Eiko
 Tsuchiya, Hiroshima University Faculty of Engineering, Department
 of Fermentation Technology, Kagami Yama, Higashi-Hiroshima,
 Hiroshima 739, Japan (E-mail: etsuchi@ipc.hiroshima-u.ac.jp,
 Tel: 0824-24-7868, Fax: 0824-24-7868).

AUTHORS Tsuchiya, E., Matsuzaki, G., Tsukao, A. and Miyakawa, T.
 TITLE Molecular cloning and characterization of S. cerevisiae HST1 gene
 JOURNAL Unpublished (1995)
 REFERENCE 3 (sites)
 Tsuchiya, E., Matsuzaki, G., Kurano, K., Fukuchi, T., Tsukao, A. and
 Miyakawa, T.
 The Saccharomyces cerevisiae SSD1 gene is involved in the tolerance
 to high concentration of Ca²⁺ with the participation of
 HST1/NRC1/BPR1
 Gene 176 (1-2), 35-38 (1996)

TITLE
 JOURNAL
 MEDLINE
 FEATURES
 source
 Location/Qualifiers
 1..4156
 /organism="Saccharomyces cerevisiae"
 /strain="RAY-3a"
 /db_xref="taxon:4932"
 946..3693
 /gene="HST1"
 946..3693
 /gene="HST1"
 /standard_name="high-copy suppressor of TRP sensitivity"
 /codon_start=1
 /protein_id="BAA08819.1"
 /db_xref="GI:1616450"
 /translation="MSNRNESHMLRTSSDKAIAISDRKRESEVILAAOISLNEIRSY
 KIKRLSIGSDLLIDPELDIRFGESGSRMSGTTSSASAPSTTYNNTRYSDP
 PLENIHGRNGSIESNKTGQNGYIGIKGVHSPERKIANVLAKNLWVPAHQHP
 VPDNLEIVORTIONIOISDNGEDNDGNSNENNDLEDNGESQSYRKNENTINL
 NGLSRHNASLIRRSRTLRSTYTERDDNDKDSASEYVKNYERLSKKEBPV
 SIRDITEELRLSNSAGLTDNDITLRLLSMAIGSDKKDQPPQPGHIDEGIGEST

THIS PAGE BLANK (1.3PT6)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 12, 2002, 09:26:44 ; Search time 17.02 Seconds

(without alignments)
975.862 Million cell updates/sec

Title: US-09-165-460A-2

Perfect score: 2342

Sequence: 1 MEDLKITLIDHPNIPKLIIS.....HPTLAERSTALDYSEKKKN 453

Scoring table: BL0SUM62

Gapop 10.0, Capext 0.5

Searched: 100059 seqs, 36664827 residues

otal number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|------------|--------------------|
| 1 | 2336 | 99.7 | 453 | ST24_YEAST | P47154 saccharomyc |
| 2 | 903 | 38.6 | 474 | ST24_SCHPO | 010071 schizosacch |
| 3 | 698.5 | 29.8 | 475 | FNCL_HUMAN | 075844 homo sapien |
| 4 | 328 | 14.0 | 426 | YHEN_BACSU | P40769 bacillus su |
| 5 | 161 | 6.9 | 286 | HTPX_PASMD | P57846 pasteurella |
| 6 | 150.5 | 6.4 | 310 | HTPX_HELPJ | 092854 helicobacte |
| 7 | 150 | 6.3 | 335 | HTPX_ARCFU | 030004 archaeoglob |
| 8 | 147.5 | 6.3 | 283 | HTPX_HAEIN | P44840 haemophilus |
| 9 | 142.5 | 6.1 | 310 | HTPX_HELPY | 025882 helicobacte |
| 10 | 139 | 5.9 | 356 | HTPX_AERPE | 09y667 aeropyrum p |
| 11 | 138.5 | 5.9 | 258 | HTPX_METHH | 026669 methanobact |
| 12 | 130.5 | 5.6 | 292 | HTPX_PYRHO | 058997 pyrococcus |
| 13 | 130.5 | 5.6 | 317 | HTPX_THEAC | 09h1v2 thermoplasm |
| 14 | 127.5 | 5.4 | 2136 | YCF2_MARPO | P09875 marchantia |
| 15 | 120 | 5.1 | 292 | HTPX_BUCAL | 025882 helicobacte |
| 16 | 119.5 | 5.1 | 2294 | YCF2_ARATH | 09y667 aeropyrum p |
| 17 | 119 | 5.1 | 297 | HTPX_STRGC | 026669 methanobact |
| 18 | 116 | 5.0 | 293 | HTPX_ECOLI | 058997 pyrococcus |
| 19 | 113 | 4.8 | 284 | HTPX_METTA | 09h1v2 thermoplasm |
| 20 | 113 | 4.8 | 284 | HTPX_METTA | 09h1v2 thermoplasm |
| 21 | 109 | 4.7 | 506 | YCX2_EUGGR | 031657 bacillus su |
| 22 | 109 | 4.7 | 595 | YCB6_METTA | P31821 euglena gra |
| 23 | 107.5 | 4.6 | 1032 | MT18_YEAST | Q58688 methanococc |
| 24 | 106.5 | 4.5 | 646 | NTPL_HAEPI | P40469 saccharomyc |
| 25 | 106.5 | 4.5 | 1029 | YFC5_YEAST | Q37371 heliothis a |
| 26 | 105.5 | 4.5 | 873 | COX1_ACACA | P43571 saccharomyc |
| 27 | 105 | 4.5 | 1427 | SRB8_YEAST | Q37370 acanthamoeb |
| 28 | 104.5 | 4.5 | 585 | YHD9_YEAST | P25648 saccharomyc |
| 29 | 104.5 | 4.5 | 1100 | LOSI_YEAST | P38732 saccharomyc |
| 30 | 104 | 4.4 | 302 | HTPX_AOUAE | P33418 saccharomyc |
| 31 | 104 | 4.4 | 530 | CP14_CHICK | 067798 aquifex aeo |
| 32 | 103.5 | 4.4 | 185 | GU01_RAT | P79690 gallus gall |
| 33 | 103.5 | 4.4 | 450 | VATH_SCHPO | P35894 ratius norv |
| | | | | | 014265 schizosacch |

ALIGNMENTS

| RESULT | 1 | ST24_YEAST | STANDARD | PRT | 453 AA |
|--------|--|------------|----------|-----|--------|
| AC | P47154 | ST24_YEAST | STANDARD | PRT | 453 AA |
| DT | 01-FEB-1996 (Rel. 33, Created) | | | | |
| DT | 01-FEB-1996 (Rel. 33, Last sequence update) | | | | |
| DT | 20-AUG-2001 (Rel. 40, Last annotation update) | | | | |
| DE | CAAX PRENYL PROTEASE 1 (EC 3.4.24.-) (PRENYL PROTEIN-SPECIFIC ENDOPEPTIDASE 1) (PPEP 1) (A-FACTOR CONVERTING ENZYME). | | | | |
| GN | STE24 OR APC1 OR YJR117H OR J2032. | | | | |
| OS | Saccharomyces cerevisiae (Baker's yeast). | | | | |
| OC | Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; | | | | |
| OC | Saccharomycetales; Saccharomycetaceae; Saccharomyces. | | | | |
| OX | NCBI_TaxID=4932; | | | | |
| RN | [1] | | | | |
| RP | SEQUENCE FROM N.A. | | | | |
| RA | Rose M., Koetter P., Entian K.D.; | | | | |
| RL | Submitted (SEP-1995) to the EMBL/Genbank/DBJ databases. | | | | |
| RN | [2] | | | | |
| RP | SEQUENCE FROM N.A., AND FUNCTION. | | | | |
| RA | MEDLINE=97167681; PubMed=90155299; | | | | |
| RA | Fujimura-Kamada K., Nouvet F.J., Michaelis S.; | | | | |
| RT | "A novel membrane-associated metalloprotease, Ste24p, is required for the first step of NH2-terminal processing of the yeast a-factor precursor."; | | | | |
| RL | J. Cell Biol. 136:271-285(1997). | | | | |
| RN | [3] | | | | |
| RP | FUNCTION. | | | | |
| RX | MEDLINE=97218305; PubMed=9065405; | | | | |
| RA | Boyartchuk V.L., Ashby M.N., Rine J.; | | | | |
| RT | "Modulation of Ras and a-factor function by carboxyl-terminal proteolysis."; | | | | |
| RL | Science 275:1796-1800(1997). | | | | |
| RN | [4] | | | | |
| RP | FUNCTION. | | | | |
| RX | MEDLINE=98365461; PubMed=9700155; | | | | |
| RA | Tam A., Nouvet F.J., Fujimura-Kamada K., Slunt H., Sisodia S.S., | | | | |
| RT | Michaelis S.; | | | | |
| RT | "Dual roles for Ste24p in yeast a-factor maturation: NH2-terminal proteolysis and COOH-terminal CAAX processing."; | | | | |
| RL | J. Cell Biol. 142:635-649(1998). | | | | |
| RN | [5] | | | | |
| RP | FUNCTION. | | | | |
| RX | MEDLINE=98393572; PubMed=9725832; | | | | |
| RA | Boyartchuk V.L., Rine J.; | | | | |
| RT | "Roles of prenyl protein proteases in maturation of Saccharomyces cerevisiae a-factor."; | | | | |
| RL | Genetics 150:95-101(1998). | | | | |
| RN | [6] | | | | |
| RP | FUNCTION. | | | | |
| RX | MEDLINE=20158928; PubMed=10692417; | | | | |
| RA | Schmidt W.K., Tam A., Michaelis S.; | | | | |
| RT | "Reconstitution of the Ste24p-dependent N-terminal proteolytic step in yeast a-factor biogenesis."; | | | | |
| RL | J. Biol. Chem. 275:6227-6233(2000). | | | | |
| RN | [7] | | | | |

| | | | | | | |
|----|-------|-----|------|---|------------|--------------------|
| 34 | 103 | 4.4 | 645 | 1 | VP74_NPVAC | P15963 autographa |
| 35 | 103 | 4.4 | 1101 | 1 | KOE5_YEAST | Q08217 saccharomyc |
| 36 | 102 | 4.4 | 577 | 1 | THH1_SCHPO | Q09684 schizosacch |
| 37 | 102 | 4.4 | 2493 | 1 | YBA4_YEAST | P35194 saccharomyc |
| 38 | 100.5 | 4.3 | 260 | 1 | YMI6_RECAM | Q12266 recitmonona |
| 39 | 100.5 | 4.3 | 1040 | 1 | RIK1_SCHPO | Q10426 schizosacch |
| 40 | 100 | 4.3 | 1325 | 1 | MRP4_HUMAN | Q15439 homo sapien |
| 41 | 99.5 | 4.2 | 309 | 1 | 07FA_HUMAN | Q76100 homo sapien |
| 42 | 99 | 4.2 | 485 | 1 | YC11_KLEBN | Q48457 klebsiella |
| 43 | 99 | 4.2 | 1024 | 1 | RPOB_PLAFA | P21421 plasmodium |
| 44 | 98.5 | 4.2 | 313 | 1 | NO2M_RHISA | Q99817 rhipicephal |
| 45 | 98.5 | 4.2 | 417 | 1 | O65A_DROME | P82982 drosophila |

CHARACTERIZATION. Pubmed-10825201;
 RA Trueblood C.E., Boyartchuk V.L., Picooglou E.A., Rozema D.,
 RA Poulter C.D., Rine J.,
 RT "The Caax proteases, Afpjp and Rcepj, have overlapping but distinct
 RT substrate specificities." Mol. Cell. Biol. 20:4381-4392(2000).
 RN [8]
 RP SUBCELLULAR LOCATION.
 RX MEDLINE-98409630; Pubmed-9736709;
 RA Schmidt W.K., Tam A., Fujimura-Kamada K., Michaelis S.;
 RT "Endoplasmic reticulum membrane localization of Rcepj and Ste24p,
 RT yeast proteases involved in carboxyl-terminal CAAX protein processing
 RT and amino-terminal a-factor cleavage." Proc. Natl. Acad. Sci. U.S.A. 95:1175-1180(1998).
 CC -1- FUNCTION: PROTEOLYTICALLY REMOVES THE C-TERMINAL THREE RESIDUES OF
 CC FARNESYLATED A-FACTOR MATING PHEROMONE. ALSO ACTS TO CLEAVE THE N-
 CC TERMINAL EXTENSION OF THE PHEROMONE. DOES NOT ACT ON RAS.
 CC -1- COFACTOR: BINDS ONE ZINC ION (PROBABLE).
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. ENDOPLASMIC
 CC RETICULUM.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M48 (ZINC
 CC METALLOPROTEASES).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: 249617; CAAB9647.1; -
 DR EMBL: U77137; AAB3827.1.1; -
 DR MEROPS: M48.001; -
 DR SGD: S0003878; STE24.
 DR InterPro: IPR001915; Peptidase_M48.
 DR InterPro: IPR000130; Zn_MTPeptide.
 DR Pfam: PF01435; Peptidase_M48; 1.
 DR PROSITE: PS00142; ZINC_PROTEASE; 1.
 KW Hydrolase: Metalloprotease; Zinc; Transmembrane;
 KW Endoplasmic reticulum; Pheromone response.
 FT TRANSMEM 13
 FT TRANSMEM 33
 FT TRANSMEM 90 110
 FT TRANSMEM 122 142
 FT TRANSMEM 168 188
 FT TRANSMEM 198 218
 FT TRANSMEM 307 327
 FT TRANSMEM 358 378
 FT METAL 297 297
 FT ACT_SITE 298 298
 FT METAL 301 301
 FT METAL 390 390
 FT ACT_SITE 394 394
 SQ SEQUENCE 453 AA; 52324 MW; 331CC9AE2D7C99DA CRC64;

Query Match 99.7%; Score 2336; DB 1; Length 453;
 Best Local Similarity 99.8%; Pred. No. 3.7e-160;
 Matches 453; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MFDTKTLIDHPNIPMKLISGFSIAQSFESYLYROYOKLSEKTLPPVLEDEIDDETFH 60
 DB 1 MFDTKTLIDHPNIPMKLISGFSIAQSFESYLYROYOKLSEKTLPPVLEDEIDDETFH 60
 QY 61 KSRNYSAPAKKFSIFGDVYMLAQKLVFIKYDLPKIMHMAVSLNAVLPAVFHNVSTVAQ 120
 DB 61 KSRNYSAPAKKFSIFGDVYMLAQKLVFIKYDLPKIMHMAVSLNAVLPAVFHNVSTVAQ 120
 QY 121 SLCEFLGLSLSTLVDPLEPSYSHFVLEKRGFKNLTVQIMTDMISLTLAVAGPIL 180
 DB 121 SLCEFLGLSLSTLVDPLEPSYSHFVLEKRGFKNLTVQIMTDMISLTLAVAGPIL 180

QY 181 YLFKIDKRPPTDPLKIMFVFLVQILAMTITVFIMPKNKPTPLEDGELKLSISLSA 240
 DB 181 YLFKIDKRPPTDPLKIMFVFLVQILAMTITVFIMPKNKPTPLEDGELKLSISLSA 240
 QY 241 DRVGFPPDKRIFIDGSRSSSHSNAYFGLPPTSRIYVFDVLVNSNDELTAVAHIG 300
 DB 241 DRVGFPPDKRIFIDGSRSSSHSNAYFGLPPTSRIYVFDVLVNSNDELTAVAHIG 300
 QY 301 HMOKNHIVNMYFISQHTFLFSLFTSIYRNTSFYNTFGFLEKSTGSGFVDPVITKEPPI 360
 DB 301 HMOKNHIVNMYFISQHTFLFSLFTSIYRNTSFYNTFGFLEKSTGSGFVDPVITKEPPI 360
 QY 361 IIGFMLENDLITPLECAMQFVMSLISRTHEYQADAVAKKIGYKNLCRALIDQIKNST 420
 DB 361 IIGFMLENDLITPLECAMQFVMSLISRTHEYQADAVAKKIGYKNLCRALIDQIKNST 420
 QY 421 MNVDLYSSYHSHPTLAERSTADYSEKKN 453
 DB 421 MNVDLYSSYHSHPTLAERSTADYSEKKN 453

RESULT 2
 ID ST24_SCHPO STANDARD; PRT; 474 AA.
 AC Q10071;
 DT 01-FEB-1996 (Rel. 33, Created)
 DT 01-FEB-1996 (Rel. 33, last sequence update)
 DT 20-AUG-2001 (Rel. 40, last annotation update)
 DE PROBABLE CAAX PRENYL PROTEASE 1 (EC 3.4.24.-) (PRENYL PROTEIN-
 DE SPECIFIC ENDOPROTEASE 1) (PPSEP 1).
 GN SPAC3H1.05.
 OS Schizosaccharomyces pombe (fission yeast).
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;
 OC Schizosaccharomyces.
 OX NCBI_TaxID=4896;
 RX [1]
 RP SEQUENCE FROM N.A.
 RC SRRAIN-972;
 RA Gentles S., Churcher C.M., Barrell B.G., Rajandream M.A., Walsh S.V.;
 RL Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: PROTEOLYTICALLY REMOVES THE C-TERMINAL THREE RESIDUES OF
 CC FARNESYLATED PROTEINS (BY SIMILARITY).
 CC -1- COFACTOR: BINDS ONE ZINC ION (POTENTIAL).
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. ENDOPLASMIC
 CC RETICULUM (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M48 (ZINC
 CC METALLOPROTEASES).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: Z68144; CA92258.1; -
 DR MEROPS: M48.001; -
 DR InterPro: IPR001915; Peptidase_M48.
 DR InterPro: IPR000130; Zn_MTPeptide.
 DR Pfam: PF01435; Peptidase_M48; 1.
 DR PROSITE: PS00142; ZINC_PROTEASE; 1.
 KW Hypothetical protein; Hydrolase; Metalloprotease; Zinc; Transmembrane;
 KW Endoplasmic reticulum.
 FT TRANSMEM 103 123
 FT TRANSMEM 196 216
 FT TRANSMEM 230 250
 FT TRANSMEM 344 364
 FT TRANSMEM 381 401
 FT METAL 332 332
 FT ACT_SITE 333 333
 FT METAL 336 336
 ZINC (CATALYTIC) (BY SIMILARITY).

OM nucleic - nucleic search, using sw model

Run on: February 12, 2002, 10:38:12 ; Search time 4622.84 Seconds

6512.739 Million cell updates/sec

Title: US-09-165-460A-1
Perfect score: 1825

Sequence: 1 acctacctttttctatct.....aaataaacgatttaacatt 1825

Scoring table: . IDENTITY_NUC

Searched: 1472140 seqs, 8248589755 residues

Total number of hits satisfying chosen parameters: 2944280

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

| | |
|------------------|--------------------|
| Post-processing: | Minimum Match 0% |
| | Maximum Match 100% |

Listing first 45 summaries

Database

| | |
|-----|----------------|
| 1: | gb_ba:* |
| 2: | gb_htg:* |
| 3: | gb_in:* |
| 4: | gb_om:* |
| 5: | gb_ov:* |
| 6: | gb_pat:* |
| 7: | gb_ph:* |
| 8: | gb_pl:* |
| 9: | gb_pr:* |
| 10: | gb_ro:* |
| 11: | gb_sts:* |
| 12: | gb_sy:* |
| 13: | gb_un:* |
| 14: | gb_vl:* |
| 15: | em_ba:* |
| 16: | em_fun:* |
| 17: | em_hum:* |
| 18: | em_in:* |
| 19: | em_om:* |
| 20: | em_or:* |
| 21: | em_ov:* |
| 22: | em_pat:* |
| 23: | em_ph:* |
| 24: | em_pl:* |
| 25: | em_ro:* |
| 26: | em_sts:* |
| 27: | em_sy:* |
| 28: | em_un:* |
| 29: | em_vl:* |
| 30: | em_hgo_hum:* |
| 31: | em_hgo_inv:* |
| 32: | em_hgo_rod:* |
| 33: | em_hgo_hum:* |
| 34: | em_hgt_rod:* |
| 35: | em_hgt_rod:* |
| 36: | em_hgt_other:* |

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB | ID | Description |
|------------|--------|-------------|--------|----|------------|--------------------------|
| 1 | 1823.4 | 99.9 | 1825 | 8 | SCJUR117W | Z49617 S.cerevisiae |
| 2 | 1553.4 | 85.1 | 1706 | 8 | SCJ077137 | U77137 Saccharomyces |
| 3 | 342 | 15.55 | 8 | 8 | SCJUR116W | Z49616 S.cerevisiae |
| 4 | 195 | 10.7 | 99360 | 8 | NCHB11N2 | AL513444 Neurospora |
| 5 | 187 | 10.2 | 36920 | 8 | SPAC3H1 | Z68144 S. pombe ch |
| 6 | 137.2 | 7.5 | 1375 | 8 | AF533722 | AF533722 Arabidopsis |
| 7 | 136 | 7.5 | 686 | 6 | A66270 | A66270 Sequence 922 |
| 8 | 136 | 7.5 | 686 | 6 | ARI5763 | ARI5763 Sequence |
| 9 | 136 | 7.5 | 686 | 6 | E66288 | E66288 Genome DNA |
| 10 | 121 | 6.6 | 935 | 8 | SCJUR118C | Z49618 S.cerevisiae |
| 11 | 117.4 | 6.4 | 1859 | 9 | AF064667 | AF064667 Homo sapiens |
| 12 | 117.4 | 6.4 | 2963 | 9 | AB015068 | AB015068 Homo sapiens |
| 13 | 117.4 | 6.4 | 2966 | 9 | HSY13834 | HSY13834 Homo sapiens |
| 14 | 117.4 | 6.4 | 2968 | 6 | E32056 | E32056 Human Afc1 |
| 15 | 114.2 | 6.3 | 1340 | 3 | AF358443 | AF358443 Physarum |
| 16 | 98.4 | 5.4 | 78592 | 3 | AC004335 | AC004335 Drosophila |
| 17 | 98.4 | 5.4 | 102413 | 2 | AC020024 | AC020024 Drosophila |
| 18 | 92.8 | 5.1 | 41327 | 3 | CBR047719 | CBR047719 Caenorhabditis |
| 19 | 91.6 | 5.0 | 40767 | 3 | CRC04412 | CRC04412 Caenorhabditis |
| 20 | 88 | 4.8 | 317511 | 1 | CJ11168X3 | Z81461 Caenorhabditis |
| 21 | 86.6 | 4.7 | 582 | 8 | CNS0186X3 | AL139076 Campylobacter |
| 22 | 81.6 | 4.5 | 660 | 8 | CNS0186P | AL119604 Botrytis |
| 23 | 72.2 | 4.0 | 2161 | 9 | AK027674 | AK027674 Botrytis |
| 24 | 66.8 | 3.7 | 1072 | 6 | AP0055334 | AP0055334 Sequence |
| 25 | 66.6 | 3.6 | 124808 | 2 | AP000652 | AP000652 Homo sapiens |
| 26 | 65.6 | 3.6 | 16488 | 1 | AE004992 | AE004992 Halobacter |
| 27 | 65 | 3.5 | 256172 | 2 | AC005139 | AC005139 Plasmodium |
| 28 | 64.4 | 3.5 | 169423 | 2 | AC069309 | AC069309 Mus musculus |
| 29 | 64.4 | 3.5 | 186306 | 2 | AC083946 | AC083946 Mus musculus |
| 30 | 63.6 | 3.5 | 1245 | 6 | AX055202 | AX055202 Sequence |
| 31 | 63.6 | 3.5 | 19715 | 1 | AE000555 | AE000555 Helicobacter |
| 32 | 63.6 | 3.5 | 234112 | 3 | PFMAL1P2 | AE013547 Plasmodium |
| 33 | 62.4 | 3.4 | 13383 | 3 | AE001374 | AE001374 Plasmodium |
| 34 | 61.6 | 3.4 | 98734 | 2 | PFMAL1P2 | AL031745 Plasmodium |
| 35 | 61.6 | 3.4 | 14943 | 1 | AE001528 | AE001528 Helicobacter |
| 36 | 59.8 | 3.3 | 12029 | 3 | AE001392 | AE001392 Plasmodium |
| 37 | 59.8 | 3.3 | 110000 | 2 | AC087899_0 | AC087899 Mus musculus |
| 38 | 59.2 | 3.2 | 114736 | 3 | PFMAL1P3 | Z98347 Plasmodium |
| 39 | 59.2 | 3.2 | 253305 | 3 | PFMAL1P7 | AL034559 Plasmodium |
| 40 | 58.8 | 3.2 | 98899 | 2 | AL060343 | AL060373 Danio rerio |
| 41 | 58.8 | 3.2 | 318221 | 2 | PFMAL13P3 | AL049184 Plasmodium |
| 42 | 58.8 | 3.2 | 318503 | 2 | AC074166 | AC074166 Mus musculus |
| 43 | 58.4 | 3.2 | 172239 | 3 | AC069587 | AC069587 Homo sapiens |
| 44 | 58.2 | 3.2 | 12029 | 2 | AE001412 | AE001412 Plasmodium |
| 45 | 58.2 | 3.2 | 187469 | 9 | AL57312 | AL57312 Human DNA |

ALIGNMENTS

| LOCUS | SCYR117W | 1825 bp | DNA | PLN | 11-AUG-1997 |
|-------------|----------|---------|-----|-----|-------------|
| SCYR117W | | | | | |
| LOCUS | | | | | |
| DEFINITION | | | | | |
| DESCRIPTION | | | | | |
| ACCESSION | | | | | |
| VERSION | | | | | |
| KEYWORDS | | | | | |
| FEATURES | | | | | |
| ORIGIN | | | | | |
| COMMENTS | | | | | |
| SCYR117W | | | | | |
| LOCUS | | | | | |
| DEFINITION | | | | | |
| DESCRIPTION | | | | | |
| ACCESSION | | | | | |
| VERSION | | | | | |
| KEYWORDS | | | | | |
| FEATURES | | | | | |
| ORIGIN | | | | | |
| COMMENTS | | | | | |
| SCYR117W | | | | | |
| LOCUS | | | | | |
| DEFINITION | | | | | |
| DESCRIPTION | | | | | |
| ACCESSION | | | | | |
| VERSION | | | | | |
| KEYWORDS | | | | | |
| FEATURES | | | | | |
| ORIGIN | | | | | |
| COMMENTS | | | | | |
| SCYR117W | | | | | |
| LOCUS | | | | | |
| DEFINITION | | | | | |
| DESCRIPTION | | | | | |
| ACCESSION | | | | | |
| VERSION | | | | | |
| KEYWORDS | | | | | |
| FEATURES | | | | | |
| ORIGIN | | | | | |
| COMMENTS | | | | | |
| SCYR117W | | | | | |
| LOCUS | | | | | |
| DEFINITION | | | | | |
| DESCRIPTION | | | | | |
| ACCESSION | | | | | |
| VERSION | | | | | |
| KEYWORDS | | | | | |
| FEATURES | | | | | |
| ORIGIN | | | | | |
| COMMENTS | | | | | |
| SCYR117W | | | | | |
| LOCUS | | | | | |
| DEFINITION | | | | | |
| DESCRIPTION | | | | | |
| ACCESSION | | | | | |
| VERSION | | | | | |
| KEYWORDS | | | | | |
| FEATURES | | | | | |
| ORIGIN | | | | | |
| COMMENTS | | | | | |
| SCYR117W | | | | | |
| LOCUS | | | | | |
| DEFINITION | | | | | |
| DESCRIPTION | | | | | |
| ACCESSION | | | | | |
| VERSION | | | | | |
| KEYWORDS | | | | | |
| FEATURES | | | | | |
| ORIGIN | | | | | |
| COMMENTS | | | | | |
| SCYR117W | | | | | |
| LOCUS | | | | | |
| DEFINITION | | | | | |
| DESCRIPTION | | | | | |
| ACCESSION | | | | | |
| VERSION | | | | | |
| KEYWORDS | | | | | |
| FEATURES | | | | | |
| ORIGIN | | | | | |
| COMMENTS | | | | | |
| SCYR117W | | | | | |
| LOCUS | | | | | |
| DEFINITION | | | | | |
| DESCRIPTION | | | | | |
| ACCESSION | | | | | |
| VERSION | | | | | |
| KEYWORDS | | | | | |
| FEATURES | | | | | |
| ORIGIN | | | | | |
| COMMENTS | | | | | |
| SCYR117W | | | | | |
| LOCUS | | | | | |
| DEFINITION | | | | | |
| DESCRIPTION | | | | | |
| ACCESSION | | | | | |
| VERSION | | | | | |
| KEYWORDS | | | | | |
| FEATURES | | | | | |
| ORIGIN | | | | | |

VERSION . 249617.1 GI:1015836

ORGANISM

REFERENCE
1 (bases 1 to 1825)

REFERENCE 2 (bases 1 to 1825)

JOURNAL

Submitted (25-SEP-1995) Data collected by MIPS on behalf of the European yeast chromosome X sequencing project. MIPS at the Max-Planck-Institut fuer Biochemie, Am Klopferspitz 18a D-82152

Oy 271 aaaccactctctaattacccggaaggaagaataaaaggaggaatagaaaactgc 330
| | | | |
Db 1 AAACCCCTCTCTAATTAAcAGGAAAAAGGAAGCAAAAAAGAGCGAATAGAAAACTC 60

| | | | |
|----|------|--|------|
| QY | 331 | aggcccttaatcaatggtttgatccttaagaacgagctctcgaccatcctcaataataccgctggaaa | 390 |
| Db | 61 | AGGCGTTTATTCATGTTGATCTTAAAGCAGATCTTCGACACATCCCTTAATAATCCCGTGAAA | 120 |
| QY | 391 | ttaacattctcgggttcgtgatgcccgaatttcttcgaaactctactctgaacgaa | 450 |
| Db | 121 | TTAAATCATTTTCGGGTTCTCGATTCCGATTCCTTAATTTCTTGAAATCTTACTGACGTACGA | 180 |
| QY | 451 | cagtaacgaagcactctcgaacaagaattgcccactgctgctgaaagcgaactgatgat | 510 |
| Db | 181 | CAGTACGAGAACCTATCTGAAACAAGTTGCCACCTGTGCTGGAAGACGAATATGATGAT | 240 |
| QY | 511 | gaacatttcataaatcaagaagaaactctccgggcccgaagccaaatctctcaattcgg | 570 |
| Db | 241 | GAACCTTTTTCATTAATAATCAAGGAAGTACTCCCGGGCCAAAGGCCAAATTTCCATTTTCGGT | 300 |
| QY | 571 | gacgcttaaacctgagcccaaaagctagtcttcatcaataaagacctctctccctaaatc | 630 |
| Db | 301 | GACGCTTAATAACCTTAGCCCCAAAAAGGTAGTTTCATCAATAATAGACCTCTTCCCTAAATC | 360 |
| QY | 631 | tggacaatggccgttctcttattgatagtcagtcgccgcagtcagaattcataatgctcc | 690 |
| Db | 361 | TGGCACAATGGCGTTTCTTATGTAATGCAAGTCCTCCGCTAGATTTATATGCTCC | 420 |
| QY | 691 | actgtgcacagaagttatcgtctctctggtgctctataccagttctgctcaactggtgat | 750 |
| Db | 421 | ACTGTGCGACAGAGTTTATGCTTCTTGGGTTCTTATCCAGTTTGCTACCTGTTGAT | 480 |
| QY | 751 | tggcaactctctactatagccaatttgccttcgggaagaaaatttggttcaataatg | 810 |
| Db | 481 | TTGCGACTCTCTTACTATTAAGCCATTTTGTCCGGAAGAAAATTTGGTTCAATAAATTG | 540 |
| QY | 811 | accgtccaactatggatcaacgaatgatgaacaagagtcacacttgccgatgctatggt | 870 |
| Db | 541 | ACCGTCCAACTATGATATCCGATATGATCAAGATCAAGATGCTACTTTGGCGTATGCTATGGT | 600 |
| QY | 871 | ggcccaatccttaccgtgtctcttaagaactcttgataaattccctactgattctccttgg | 930 |
| Db | 601 | GGCCCAATCTTACTCTGTTCCCTTAAGATCTTTGATAAATTTCCCTACTATTTCTTTGG | 660 |
| QY | 931 | tacattatggtttcttctgtgtgtgccaactcttagccatgacaatacttcagcttc | 990 |
| Db | 661 | TACATTATGTTCTTCTGTGCTGTGTCCAAATCTTAGCCATGACAATTCACAGCTTC | 720 |
| QY | 991 | atcatgcccacttttaataagttcacctccatvgaaagcgtgaactggaataaatctatc | 1050 |
| Db | 721 | ATCATGCCCATGTTTAATAAGTTACATCCTCATTTGGAGAGCGTCACTGAAAAATCTATT | 780 |
| QY | 1051 | gaaagtgtgcgataagatgtgggtccctctagataaagatttgtloaavgcgtca | 1110 |
| Db | 781 | GAAGATTGTGGCGATGAGATTTGGGTCCCTTAGATAAATTTTGTCAATTGACGGCTCA | 840 |
| QY | 1111 | aaaagctcttcatcaagaacgatacttcaagaatttcccaatcttaccctccaagaaat | 1170 |
| Db | 841 | AAAAGATCTTCTCATTCACAAAGCATATTTTCACAGGTTTCCATTCACCTCCAAAGAAAT | 900 |
| QY | 1171 | gtttgttcgacaactttagtgaacaglaattctacgatvgaaatbaacggctgtttggcc | 1230 |
| Db | 901 | GTTTGTTCGACACTTTAGTGAACAGTAAATTCTACTGATGAATATTCGGCTGTTTGGCC | 960 |
| QY | 1231 | catgaaatcgtgcaactgagcaaaaaaacacacatcgtaataatgltcaactttagtcaatg | 1290 |
| Db | 961 | CATGAATATGGGTACGTGCGCAAAAAACCAACATCGTTAATAATGTCTACTTTATCTAATG | 1020 |
| QY | 1291 | caaaccttctcaattctccaccttcacacagatctacagaataatcatcatcttaaac | 1350 |
| Db | 1021 | CACACCTTCTCATTTTCTCCCTTTACACAGCATTTACAGAAATACATCATATTTACAAC | 1080 |
| QY | 1351 | acccttgcgcttttctttagaagaagtcacactgagcttgttgcataccgttaactaaag | 1410 |
| Db | 1081 | ACCTTGCGGCTTTTCTTAGGAAGTCCACTGGGAGTTTGTGTATCCCGTTATCACTAAG | 1140 |
| QY | 1411 | gaattcccaatatcatgtgattatglttaatttaacgacttaatactcaactcgaaagt | 1470 |

THIS PAGE BLANK (1.3PT0)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 12, 2002, 10:38:12 ; Search time 4622.84 Seconds
(without alignments)
6512.739 Million cell updates/sec

Title: US-09-165-460A-1
Perfect score: 1825
Sequence: 1 acctacattttttctatct.....aaataaacgattaacatt 1825

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : GenEmbl:
1: gb_ba:*
2: gb_hcg:*
3: gb_in:*
4: gb_cm:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_cm:*
20: em_or:*
21: em_ov:*
22: em_pat:*
23: em_ph:*
24: em_pl:*
25: em_ro:*
26: em_sts:*
27: em_sy:*
28: em_un:*
29: em_vl:*
30: em_hcgo_hum:*
31: em_hcgo_inv:*
32: em_hcgo_rnd:*
33: em_hcg_hum:*
34: em_hcg_inv:*
35: em_hcg_rnd:*
36: em_hcg_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|--------|-------------|--------|-------|--------------------|
| 1 | 1823.4 | 99.9 | 1825 | 8 | SCYJR117W |
| 2 | 1553.4 | 85.1 | 1706 | 8 | SCU77137 |
| 3 | 342 | 18.7 | 1555 | 8 | SCYJR116W |
| 4 | 195 | 10.7 | 99360 | 8 | NCB11N2 |
| 5 | 187 | 10.2 | 36920 | 8 | SPAC3H1 |
| 6 | 137.2 | 7.5 | 1275 | 8 | AF353722 |
| 7 | 136 | 7.5 | 686 | 6 | AB6270 |
| 8 | 136 | 7.5 | 686 | 6 | AR155763 |
| 9 | 136 | 7.5 | 686 | 6 | B66288 |
| 10 | 121 | 6.6 | 935 | 8 | SCYJR118C |
| 11 | 117.4 | 6.4 | 1859 | 9 | AF064867 |
| 12 | 117.4 | 6.4 | 2963 | 9 | AB016068 |
| 13 | 117.4 | 6.4 | 2966 | 9 | HSY13834 |
| 14 | 117.4 | 6.4 | 2966 | 6 | E32056 |
| 15 | 114.2 | 6.3 | 1340 | 3 | AF358443 |
| 16 | 98.4 | 5.4 | 78592 | 3 | AC004335 |
| 17 | 98.4 | 5.4 | 102413 | 2 | AC020024 |
| 18 | 92.8 | 5.1 | 41327 | 3 | CBRG47319 |
| 19 | 91.6 | 5.0 | 40767 | 3 | CEC04F12 |
| 20 | 88 | 4.8 | 317511 | 1 | CJ11168X3 |
| 21 | 86.6 | 4.7 | 582 | 8 | CNS018P |
| 22 | 81.6 | 4.5 | 660 | 9 | AK027874 |
| 23 | 72.2 | 4.0 | 2161 | 6 | AX055334 |
| 24 | 66.8 | 3.7 | 1072 | 6 | AP000652 |
| 25 | 66.6 | 3.6 | 124808 | 2 | AE004992 |
| 26 | 65.6 | 3.6 | 16488 | 1 | AE004992 |
| 27 | 65 | 3.6 | 256172 | 2 | AC005139 |
| 28 | 64.4 | 3.5 | 169423 | 2 | AC069309 |
| 29 | 64.4 | 3.5 | 186306 | 2 | AC083946 |
| 30 | 63.6 | 3.5 | 1245 | 6 | AX055202 |
| 31 | 63.6 | 3.5 | 19715 | 1 | AE000555 |
| 32 | 63.6 | 3.5 | 234112 | 3 | PFMAL4P2 |
| 33 | 62.4 | 3.4 | 13383 | 3 | AE001374 |
| 34 | 61.8 | 3.4 | 98734 | 2 | PFMAL1P2 |
| 35 | 61.6 | 3.4 | 14943 | 1 | AE001528 |
| 36 | 59.8 | 3.3 | 12029 | 3 | AE001392 |
| 37 | 59.8 | 3.3 | 110000 | 2 | AC087899.0 |
| 38 | 59.2 | 3.2 | 114736 | 3 | PFMAL3P3 |
| 39 | 59.2 | 3.2 | 253305 | 3 | PFMAL3P7 |
| 40 | 58.8 | 3.2 | 98899 | 2 | AL603743 |
| 41 | 58.8 | 3.2 | 318221 | 2 | PFMAL1P3 |
| 42 | 58.8 | 3.2 | 318503 | 2 | AC074166 |
| 43 | 58.4 | 3.2 | 172239 | 2 | AC069587 |
| 44 | 58.2 | 3.2 | 112029 | 3 | AE001412 |
| 45 | 58.2 | 3.2 | 187469 | 9 | AL357312 Human DNA |

ALIGNMENTS

| RESULT | 1 | LOCUS | SCYJR117W | 1825 bp | DNA | PLN | 11-AUG-1997 |
|------------|---|-------|-----------|---------|-----|-----|--|
| DEFINITION | | | | | | | S.cerevisiae chromosome X reading frame ORF YJR117W. |
| ACCESSION | | | | | | | Z49617.Y13136 |
| VERSION | | | | | | | Z49617.1 GI:1015836 |
| KEYWORDS | | | | | | | |
| SOURCE | | | | | | | |
| ORGANISM | | | | | | | Baker's yeast. |
| | | | | | | | Saccharomyces cerevisiae |
| | | | | | | | Eukaryotic; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; |
| | | | | | | | Saccharomycetaceae; Saccharomycetaceae; Saccharomycetes. |
| REFERENCE | | | | | | | 1 (bases 1 to 1825) |
| AUTHORS | | | | | | | Rose, M., Koelter, P. and Entlian, K.D. |
| JOURNAL | | | | | | | Unpublished |
| REFERENCE | | | | | | | 2 (bases 1 to 1825) |
| AUTHORS | | | | | | | MIPS. |
| TITLE | | | | | | | Direct Submission |
| JOURNAL | | | | | | | Submitted (25-SEP-1995) Data collected by MIPS on behalf of the European yeast chromosome X sequencing project. MIPS at the Max-Planck-Institut fuer Biochemie, Am Klopfersplatz 18a D-82152 |

Marlinried, FRG: E-mail: Mewes@mps.embnet.org

FEATURES
sourceLocation/Qualifiers
1. 1825

/organism="Saccharomyces cerevisiae"

/db_xref="taxon:4932"

/chromosome="X"

343. 1704

/gene="STE24"

343. 1704

/note="ORF YOR117w"

/protein_id="CA89647.1"

/db_xref="GI:1015837"

/db_xref="SWISS-PROT:P47154"

/translaction="MFDLKITLDHPNIPKMLIISGFSIADGFSFSLTYRQXKLSF
KLPVLEDEIDDETFHKSRLNRAKAESEJEDVYNIAOKLVEFKYDLFPKIMHAYS
LNNVLEPRHNVSTVAOSLCFLGLISLVLPLPSYSHFVLEERKFGNKLTVOL
WITDKISLILAVYIGPILYFLKIPDKPPTDLMVLELVVOILAMTIPVFLM
PMFNKTPLEDGELKSIESIADRVGFLPKIFVIDSKSSHSNATFQGLIPFMSKRI
VLDDTLVNSDDEITRAVLAHEIGHOKNHLVNNVITSQHLTFILFSLFSTIYNISF
YNTGFELEKSTGSEFVDVPTKEPPIIIGFVLENDLITPLECAMQFVMSLISFHEHQ
ADAVAKKLGKONICRALIDLOIKNLSTMVNDPLYSYHSHPTLRRLTALDYSEK
KKN"

BASE COUNT 570 a 397 c 280 g 578 t

ORIGIN

Query Match

99.9% Score 1823.4; DB 8; Length 1825;

Best Local Similarity 99.9%; Pred. No. 0; Matches 1824; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 accatcctttttctatctcaacaagaaagccttaacacacacacacacacatc 60
DB 1 ACCATCCTTTTCTATCTCAACACGAAAGCGCTTACACACACACATCATCT 60
QY 61 acatcacatcacatataatataatagtaactgtatattctctcttaaccaa 120
DB 61 ACATCACATCACATAAATATACATATATGTAATCTGATCTCTATTAACCAA 120
QY 121 aagaagcaatataactttccctctttcttaagtcattacttaacaaactta 180
DB 121 AAGAGGCAATTAACCTTTTCCCTCTTTTACGTCATTACTCAAAACCTTA 180
QY 181 tcgtctctgtctgcacatttcccgagaataatcgacgggaataaataaagac 240
DB 181 TCGTCTCTGTCTGCCATTTCCTCAGAAAAAATCGAGGGAATTAATAAAA 240
QY 241 aacgaacaagaagaagaagtcggaatataaacacactcttaataacagga 300
DB 241 AACGACACAGAGAAAGTTGCGAATATTAACCCCTCTATTAATTAACAG 300
QY 301 aggaataaagaagaagaataagaaactgcagagccttattcattgttgatc 360
DB 301 AGGAAAAAAGAGGAATAGAAATCGAGGCTTATTCACTTTGATCTTAAGC 360
QY 361 attctgcacacttaataatccgtggaataatcattctgggtctcgattgcca 420
DB 361 ATTCTGCACACTCTAATATCCCGTGAATTAATCATTTCTGGGTTCTGAT 420
QY 421 tttctctggaatcttcttgagtcagacagatcaggaagtcatactcgaaaca 480
DB 421 TTTCTCTTTCGATTTCTTGACGTACAGACAGTACAGAACTATCTGAACA 480
QY 481 caacctgtctgagaagcgaatgatgaatgaactttcataataagaactactcc 540
DB 481 CCACCTTGCTGGAAGACGAATGATGATGAACCTTTCTAATAATCAAGAAC 540
QY 541 cgggccaagcgaagtcctcattctcgtagctataacccagcccaagctagt 600
DB 541 CGGGCCAAAGGCCAAGTCTTCGATTTTCGGTACGTCATTAACGCAAAAG 600
QY 601 ttcataaataagacactcttcctaaatctgacatgagccgtttcttattga 660

DB 601 TTCACTAATATCGACCTTTCCTTAATAATGCGACATGCGGCTTCTTATTAATGA 660
QY 661 gtccgcagcagatcattacataatgtcccaactgtgcagagattatgtcttcgtg 720
DB 661 GTCGCGCCAGTCAGATTTGATATGATGTCCTCCACTGCGGACAGAGATTATGCTTGGGT 720
QY 721 ccttatacagctgtcttaactctgtgattgtgcaactctcttactataagcaattgtc 780
DB 721 CTTTATCCAGATTTCTGATCTGATGATGATGATGATGATGATGATGATGATG 780
QY 781 ctggaagaagaatttggtttcaataaatttgccgttcacatgataccgattgtac 840
DB 781 CTGGAGAGAAATTTGGTTCAATTAATTAATTAATTAATTAATTAATTAATTA 840
QY 841 aagagctcgaatttgccgtatgcattgtgtgccaatcccttaactcttccctaagtc 900
DB 841 AAGAGTCGATTTGGGCTATGATGATGATGATGATGATGATGATGATGATG 900
QY 901 ttgataaattccctactgatttcccttggtatcattatgctctctgtgtccaa 960
DB 901 TTTGATTAATTCCTTACGATTTCTGATGATGATGATGATGATGATGATGAT 960
QY 961 attctaacatgacaatcattccagctctctacatgccaagtttaaatgataccca 1020
DB 961 ATCTTAGCATGACAAATCATTCAGCTTCATCATGCCCCATGTTAATTAATCA 1020
QY 1021 ttggaagcagctgtaactgtaaaatctataaagtttgccgataagtttggtccct 1080
DB 1021 TTGGAGAGCGGTGAACTGAAAATTCATTAAGTTTGGCCGATAGAGTTGGTCCCT 1080
QY 1081 ctagaataagattttgtcattgacggtcgaagaatctctcattcaacagcatattc 1140
DB 1081 CTAGATTAAGATTTTGTGATGATGACGGCTCAAAAGATCTTCATTAACGATATTC 1140
QY 1141 acaagtttgccattccactcccaagaagattgtttgttcgacacttggtgaacgtat 1200
DB 1141 ACAGGTTGGCCATTACCTCCACAGAGAAATTTGTTGTCGACACTTGTGTAAC 1200
QY 1201 tctatgataagaaactaagcgtgttttgcccaatgaatcgttcaactgcgaataaagac 1260
DB 1201 TCTATGATGAAATTAAGCGCTGTTTGGCCATTAATGCTACATGCAAAAAACAC 1260
QY 1261 atcgttaataatgatacttcttaagcaatctgcacacttccctattctcccttacc 1320
DB 1261 ATCGTTAATATGCGATCTTATGCAATGACACCTTCATTTCTCCCTTTCAAC 1320
QY 1321 agcatctacagaataatataatcttaaacacacttgcgctttctctagaagaagccact 1380
DB 1321 AGCATCTACAGAAATTAATCAATTTTACACACCTTCGCTTTCTTGAAGAA 1380
QY 1381 ggcagttctgtatcccgatatacactaagaagattcccaacttaacattggaatttga 1440
DB 1381 GCGAGTTTGTGATGCCGTTATACATTAAGAAATCCCATTAATGATGATTA 1440
QY 1441 tttaagcaacttaatactcaactcgaatgtgtcgaatcgtgatagtttaaattcc 1500
DB 1441 TTTAAGCACTTAATTAATCCGATCGAATGTCGATGCAATTCGATGATTA 1500
QY 1501 agaaactatgaataatcaagctgtatgtcttaataaataatgggctacagaacaaactca 1560
DB 1501 AGAATCATGAAATTAATCAAGCTGATGCTTAATAAATGGGCTACAAACAA 1560
QY 1561 tgttaggctcttaattgtatcacaataaataaacttccacacagaaatgataccctatg 1620
DB 1561 TGTAGGCTCTTAATTAATTAATCAATTAATTAATTAATTAATTAATTAAT 1620
QY 1621 tattctagctatcatttcccatcacaactcagctgaagaatgacgcgtctagactat 1680
DB 1621 TATTCTAGCTATCTTATTCCTATCAACTCTAGCTTAAGATTAAGCGCTCTG 1680
QY 1681 gttatgtaaaagaagaataaataatctatagagtagacatatattgacatgtacgttaaat 1740

| | | | |
|----|------|--|------|
| OY | 331 | aggcccttatcaatglttgatctctaagacgatctccaccatcccaatatcccyggaaa | 390 |
| Db | 61 | AGGCCCTTATTCATGTTTGATCTTAAGACGATTCTCGACCATCTCTAATATCCCGGGAAA | 120 |
| OY | 391 | ttatacatcttcgtgtcttcgatctgcgcacaattctcttgaactctactgaagcaga | 450 |
| Db | 121 | TTAATCATTTCTGGGTTCTCGATTGGCCCAATTTCTTTGGAATCTTACTTGACGTACGA | 180 |
| OY | 451 | cagtaacagaagatctcttgaaacaaagtctgcacactgtctgtyaagaagaaatlgatg | 510 |
| Db | 181 | CAGTACACGAAGCTATCTGAAACAAGATTGCCACCTGTGCTGTGGAAGCAATTCATGAT | 240 |
| OY | 511 | gaaactttcataaatacaagaaactactcccgggccaagycgaagtctccatlttgyt | 570 |
| Db | 241 | GAAACTTTTCAATAATCAAGAACTACTCCCGGGCCAAAGCCAAAGTCTCCATTTTCGGT | 300 |
| OY | 571 | gaactcataaccctagggccaaagaagcttagtttcatcaataaataagactcttccctaaac | 630 |
| Db | 301 | GACGCTATTAACCTTAGCCCAAAAGCTTAGTTTATATAAATAGCACTCTTCCCTAAAATC | 360 |
| OY | 631 | tggcacatgagccgtttctcttattgaaatgcagtcctgcagtcagatatctcatatgctcc | 690 |
| Db | 361 | TGGCACAAGGCCGTTTCTTATTAAGAAATGACGCTCCGCCATGAGATTATATGTGTCAC | 420 |
| OY | 691 | actgtgcagacagagtttaagctcttcgggtctctatccaagtttgcataccttgatg | 750 |
| Db | 421 | ACTGTGCGACAGAGTTTATATCTCTTGCGGCTCTTATCCAGTTTGCTACCTTTGGTTGAT | 480 |
| OY | 751 | ttgcacactctctactatagccaatttgcctcggaagaaataattggtttcaataaatg | 810 |
| Db | 481 | TTGCACTCTCTTACTATAGCCATTTTGTCTTGAAAGAAATTTGGTTTCATATAATTG | 540 |
| OY | 811 | accgtlccaactatlgatcacccgaatalgatcaagatctgactttggcgatgactatlg | 870 |
| Db | 541 | ACCGTCCAACTATGATGATCACCGAATATGATCAAGATCTGACTTGGCGATGCTATGTGT | 600 |
| OY | 871 | ggcccaatccttaactccgttctcttaagatctttgataaataatccctactgatatctcc | 930 |
| Db | 601 | GGCCCCAAATCTTTACCGTTCCTTAAATCTTTGAATAATCCCTACATATTTCCCTTGG | 660 |
| OY | 931 | tacattatgagctctctgtctcgattgtgccaaacttagccatgacaaatactccagcttc | 990 |
| Db | 661 | TACATTAATGATCTTCTTGTTCTGTTGCCAATTTTAGCCATGACAAATCTTCCAGTCTTC | 720 |
| OY | 991 | atcatgcccaatglttaataagatctcaactctgagagacggtgaactgtaaaaaactat | 1050 |
| Db | 721 | ATCATAGCCCAATGTTAATTAATGATTCACGCCATGTGAGAGAGCGTGAACGTAAAAATCTATT | 780 |
| OY | 1051 | gaaagtctggccgagatagttgggtctccctctagataaagattttgtatgtgaaggtca | 1110 |
| Db | 781 | GAAAGTTTGGCGGATGATAAGTTGGGTTCCCTCTGTGAATAAGATTTTGTCTATACGCGTCA | 840 |
| OY | 1111 | aaaagatcttctcaataaagcgaatcttaagaagtttgcactcaactccccaagaagaat | 1170 |
| Db | 841 | AAAAGATCTTCTCTTAACAGCATTAATTAACAGGTTTGCCATTACCTCCAAAGGAATT | 900 |
| OY | 1171 | gtttgttcgcagacatttagtgaacagaaatctcaacgatgaatlaacgggtcttctggcc | 1230 |
| Db | 901 | GTTTGTTCGCACCTTAAGTAACAGTAATTCTACGATGAATAATTACGCGCTGTTTGGCC | 960 |
| OY | 1231 | catgaataatcgtctactcgtgcaaaaaaacacatctgttaaatatggtcatctttagtcaatg | 1290 |
| Db | 961 | CATGAATAACGGTCACTAGGCAAAAAAACACATGTTAATAATGTCTCTTTAGTCAATTG | 1020 |
| OY | 1291 | cacacactcccatcttctccctttcacagagatctacaagaataacatcatcttacaac | 1350 |
| Db | 1021 | CACACCTTCCTCATTTTCTCCCTTTTCCACAGCATCTACAGAAATACATATTTACAAAC | 1080 |
| OY | 1351 | accttcggcttttctctagaagaatccactcgtgcagitttgttgatccgttacaactaag | 1410 |
| Db | 1081 | ACCTTCGGCTTTTCTTAGAAGAGTCCACAGCGAGTTTGTGATCCCTTATCCTACTAAG | 1140 |
| OY | 1411 | gaattcccatatcatctgattatgatttataagacttataactccactcgaagt | 1470 |

| | | | |
|----|------|--|------|
| QY | 331 | aggccttatcatcgttctgatctcttaagaagatcttcgcacatcctcaataccgtggaaa | 390 |
| Db | 61 | AGGCGTTTATTCATGTTTATCTTAAGAGATTCGACATCTCAATATCCCTGGAAA | 120 |
| QY | 391 | ttaactatctctgggttcgcagatcgccaaattctctccgaactcttaactgaagtaaga | 450 |
| Db | 121 | TTAATCTTTCTGGGTTCGATTCGCCCAATTTTCTTCGAAATCTTACTTACGTACAGA | 180 |
| QY | 451 | cagtaaccagaagcatctgaanaaagctgtgcaacctgtgctggaagacgaatgtatgat | 510 |
| Db | 181 | CAGTACAGAAAGCATCTGAAACAAAGTGGCCACCTGTCGTCGGAAGACGAATGTATGAT | 240 |
| QY | 511 | gaaactttcataatcgaagaagactactccggcgccgaagccaagttcttcatttgggt | 570 |
| Db | 241 | GAACCTTTTATTAATCAACGAAGACTCTCCCGGGCAAGGGCCAACTTCTCATTTTGGGT | 300 |
| QY | 571 | gagctcctaacctagcccaaaaagctagcttttcacccaataagacacttccctcaaaac | 630 |
| Db | 301 | GAGCTCTATTAACCTAGCCCAAAAGCTAGTTTTCATCANAATACGACCTTCCCTTAATTC | 360 |
| QY | 631 | ttgacaca ttggccgcttctctatctgaa tgcagctccgtgcagtcaga ttccalatgtctcc | 690 |
| Db | 361 | TGGCACATGGCCGTTTCTTATTAATGAATGCAAGTCGACGTCAGATTCATATGTGCTCC | 420 |
| QY | 691 | acgtctgcaagaagttatagctctctgggtctctcatccagttcttcaacttggttgat | 750 |
| Db | 421 | ACTGTCCACAGAGATTTATGCTTTCCTTGGGTCTCTTATCCAGTTTCTACCTTGGTTGAT | 480 |
| QY | 751 | ttgcacactccttaactaactaagcaattctgcgcggagaagaaaatttggtttcaataatg | 810 |
| Db | 481 | TTGGCACTCTTACATTAACCATTTTGTCTCGAAGAAAATTTGGTTCAATTAATG | 540 |
| QY | 811 | accgtccaactatgatalcaacgga talgatcaagagctctgactttggcgtatgctatgt | 870 |
| Db | 541 | ACCGTCAACATAGATGACACCGATATGATCAAGAGCTGACTTGGCGATGCTATTGCT | 600 |
| QY | 871 | ggcccaactccttaactctgtcttcctaagaactcttgataaaatccctactgattcccttgg | 930 |
| Db | 601 | GGCCCAATCCTTAACCGTTCCTTAAGATCTTGGAATTAATCCCACTACGATTTCCCTTGG | 660 |
| QY | 931 | tacatctgtctctctgtctgtctgtgcgtgcacactctgaacttgcga tgaacatctccagcttc | 990 |
| Db | 661 | TACATTTATGGTCTTCTTGTTCGTTGTCNAATCTTAGCCATACATCATTTCCACTCTTC | 720 |
| QY | 991 | atcatgcacatglttlaataaagttcactccatctggagagacggcgtgaactgaanaactcat | 1050 |
| Db | 721 | ATCATGGCCATGTTTATTAAGTCACTCCACTTGGAGAGAGGTAAGTGAANAATCTATT | 780 |
| QY | 1051 | gaaagttctggccgatagagttgggttcctctagataaagatttctgcaatgtagcgctca | 1110 |
| Db | 781 | GAAAGTTTGGCCGATAGAGTTGGGTTCCTCTTAGATTAAGATTTTGTCAATGACGGCTCA | 840 |
| QY | 1111 | aaaagatcttctatccaagaacgatcttccaaggttgcacttaactccaagagaatt | 1170 |
| Db | 841 | AAAAGATCTTCTCTTAAGCGCATTTTACAGGTTTGCCATTAACCTCCAAAGGAATTT | 900 |
| QY | 1171 | gttttgtctgacactttagtgaacaglaactctacgtatgaataactaagcgtcttggcc | 1230 |
| Db | 901 | GTTTGTTCGCACTTTAGTGAACAGATTAFTTACTAGATGAATTAACGGCTGTTTGGCC | 960 |
| QY | 1231 | catgaaactcgtctacgtggcaaaaaaacacacatcgltataa tggltacatctttagtcaatg | 1290 |
| Db | 961 | CATGAATCGGTACTCTGCAAAAAAACACATCGTTAAATATGTCATCTTTAGTCAATTTG | 1020 |
| QY | 1291 | cagacctctcctaatttctcccttccacacagatctcaagaatataactcaatttacaac | 1350 |
| Db | 1021 | CACACCTTCCCATTTCTCCCTTTTCCACGAGATCTACAGAAATACATCATTTTACAC | 1080 |
| QY | 1351 | acctctggctttctcttagaagaagtcacatcggcaggttctgtgatcccglttacaag | 1410 |
| Db | 1081 | ACCTTCGGCTTTTCTTAAAGAAAGTCCACTGGCAGTGTGTTTGATGCCGTTATCTCTAAG | 1140 |
| QY | 1411 | gaattcccatatcatctgattatgattatctaagactiatlaactcaactcgtgaagt | 1470 |

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 12, 2002, 13:20:11 ; Search time 4622.84 Seconds

(without alignments)
10520.305 Million cell updates/sec

Title: US-09-165-460A-3

Perfect score: 2948
Sequence: 1 tgaactgttgatgaacaag.....ggggaggaataaagaatcaca 2948

Scoring table:
IDENTITY_NUC
Gapex 10.0 , Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

Listing first 45 summaries

```

GenEmbl:*
1: gb_da:*
2: gb_hcg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_om:*
20: em_or:*
21: em_ov:*
22: em_pat:*
23: em_ph:*
24: em_pl:*
25: em_ro:*
26: em_sts:*
27: em_sy:*
28: em_un:*
29: em_vl:*
30: em_htgo_hum:*
31: em_htgo_inv:*
32: em_htgo_rtd:*
33: em_htg_hum:*
34: em_htg_inv:*
35: em_htg_rtd:*
36: em_htg_other:*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|--------|-------------|--------|--------------|---------------------|
| C 1 | 2948 | 100.0 | 29987 | 8 SC8156 | 249260 S.cerevisiae |
| 2 | 1560.4 | 52.9 | 4156 | 8 D50276 | D50276 Saccharomyc |
| 3 | 1361 | 46.2 | 4130 | 8 YSCNCR1G | 142821 Saccharomyc |
| 4 | 1163 | 39.5 | 4225 | 8 SCU65682 | U65682 Saccharomyc |
| 5 | 1069.4 | 36.3 | 3250 | 8 SCU22580 | U22580 Saccharomyc |
| 6 | 931 | 31.6 | 3525 | 8 SCUD3849 | U63849 Saccharomyc |
| 7 | 786.8 | 26.7 | 3240 | 8 YSCBU1 | Y80083 Yeast BTL1 |
| 8 | 703.8 | 23.9 | 3020 | 8 SCROS1GEN | X8901 S.cerevisia |
| 9 | 725.2 | 7.6 | 270 | 8 YSCGNC58X | X87379 Yeast Eco R |
| 10 | 143.6 | 4.9 | 27559 | 8 SC8339 | 249210 S.cerevisia |
| C 11 | 75 | 2.5 | 2033 | 8 SCYNR066C | Z71683 S.cerevisia |
| C 12 | 71.8 | 2.4 | 4775 | 8 SCU32938 | U32938 Saccharomyc |
| C 13 | 68 | 2.3 | 780 | 11 CNS06K8B | AL402561 T7 end of |
| C 14 | 65.6 | 2.2 | 256172 | 2 AC005139 | AC005140 Plasmodiu |
| C 15 | 65.6 | 2.2 | 310779 | 2 AC005140 | AC005140 Plasmodiu |
| C 16 | 64.4 | 2.2 | 904 | 11 CNS06189 | AL399967 T3 end of |
| C 17 | 64.4 | 2.2 | 86827 | 3 PFMA13P5 | AL034556 Plasmodiu |
| C 18 | 63.4 | 2.2 | 1920 | 3 PFCTUB | X62393 P.falciparu |
| C 19 | 63.4 | 2.2 | 110000 | 2 PFMA13P2_0 | AL049185 Plasmodiu |
| C 20 | 62 | 2.1 | 253305 | 3 PFMA13P7 | AL034559 Plasmodiu |
| C 21 | 61.6 | 2.1 | 1141 | 6 AX083744 | AX083744 Sequence |
| C 22 | 60.8 | 2.1 | 224112 | 3 PFMA14P2 | AL035475 Plasmodiu |
| C 23 | 60.6 | 2.1 | 202872 | 2 AC016160 | AC016160 Homo sapi |
| C 24 | 60 | 2.0 | 5632 | 3 AF010892 | AF010892 Dictyoste |
| C 25 | 59.4 | 2.0 | 3214 | 3 AF019980 | AF019980 Dictyoste |
| C 26 | 59.4 | 2.0 | 318221 | 2 PFMA13P3 | AL049184 Plasmodiu |
| C 27 | 58.8 | 2.0 | 321003 | 2 PFMA14P3 | AL035476 Plasmodiu |
| C 28 | 58.4 | 2.0 | 78439 | 17 HST085197 | U85197 Homo sapien |
| C 29 | 58.4 | 2.0 | 169546 | 2 AC004157 | AC004157 Plasmodiu |
| C 30 | 58.4 | 2.0 | 245802 | 2 AC006279 | AC006279 Homo sapi |
| C 31 | 58.4 | 2.0 | 251124 | 9 HUAE000660 | AE000660 Homo sapi |
| C 32 | 58.2 | 2.0 | 153098 | 3 PFMA13P2 | AL034558 Plasmodiu |
| C 33 | 58 | 2.0 | 169794 | 2 AC004688 | AC004688 Plasmodiu |
| C 34 | 58 | 2.0 | 178137 | 9 AC011302 | AC011302 Homo sapi |
| C 35 | 57.8 | 2.0 | 13684 | 3 AE001403 | AE001403 Plasmodiu |
| C 36 | 57.8 | 2.0 | 153477 | 2 AC006278 | AC006278 Plasmodiu |
| C 37 | 57.6 | 2.0 | 53932 | 2 AC023371 | AE001431 Homo sapi |
| C 38 | 57 | 1.9 | 12029 | 3 AE001431 | AE001431 Plasmodiu |
| C 39 | 57 | 1.9 | 163443 | 2 AC006280 | AC006280 Plasmodiu |
| C 40 | 57 | 1.9 | 18273 | 2 AC005308 | AC005308 Plasmodiu |
| C 41 | 57 | 1.9 | 196149 | 2 AC004709 | AC004709 Plasmodiu |
| C 42 | 56.8 | 1.9 | 67970 | 3 PFMA11P3 | AL031746 Plasmodiu |
| C 43 | 56.8 | 1.9 | 149627 | 9 AC087428 | AC087428 Homo sapi |
| C 44 | 56.6 | 1.9 | 92633 | 2 PFMA14P1_3 | Continuation (4 of |
| C 45 | 56.6 | 1.9 | 153477 | 2 AC006278 | AC006278 Plasmodiu |

ALIGNMENTS

| RESULT | 1 |
|------------|--|
| SC8156/c | |
| LOCUS | 29987 bp DNA |
| DEFINITION | S.cerevisiae chromosome XIII cosmid 8156. |
| ACCESSION | 249260.1 |
| VERSION | 249260.1 GI:809081 |
| KEYWORDS | BTL1; cytochrome b5; DAG1; initiation factor 1A; inorganic pyrophosphatase; IIP2; nitrate reductase; orotate phosphoribosyltransferase; tau element; TIF1; TPS3; transfer RNA-Gln; trehalose-phosphate synthase; U6 snRNP; URA10; baker's yeast. |
| SOURCE | Saccharomyces cerevisiae |
| ORGANISM | Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; Saccharomycetales; Saccharomycetaceae; Saccharomyces. |
| REFERENCE | 1 (bases 1 to 29987) |
| AUTHORS | Iye,G., and Churcher,C.M. |
| JOURNAL | Unpublished |
| REFERENCE | 2 (bases 1 to 29987) |
| AUTHORS | Barrell,B. and Rajandream,M.A. |
| TITLE | Direct Submission |

JOURNAL

COMMENT

Submitted (12-MAY-1995) *Saccharomyces cerevisiae* chromosome XIII sequencing project, Sanger Centre, Hinxton Hall, Hinxton, Cambridge CB10 1HQ E-mail: barrell@sanger.ac.uk

Notes:

All CDS over 100 codons have been analysed. CDS that are completely overlapped and those that are overlapped by more than 50% of their length by a larger CDS have been omitted from this analysis. Details of the omitted CDS are available on request. The more significant matches with motifs in the PROSITE database are also included but some of these may be fortuitous. The length in codons and the calculated codon adaptation index (CAI) is given for each CDS. Cosmid 8156 is overlapped at the start by cosmid 9920, emb1 entry SC9920, accession no. Z46639 and at the end by cosmid 8021.

FEATURES

source

1. 29987

/organism="Saccharomyces cerevisiae"

/strain="AB972"

/db_xref="taxon:4932"

/chromosome="XIII"

/clone="cosmid 8156"

/map="13R"

/complement(<1..356)

/note="YM8156.01c, unknown, len: 118, CAI: 0.14, overlaps and extends YM9920.13c"

/codon_start=1

/product="unknown"

/protein_id="CA89242.1"

/db_xref="GI:809082"

/db_xref="SWISS-PROT:003496"

/translation="MTDVOFSSDEIEFRVKEFLIANNPANNENDAVLQIEHD FRYLIQYKDGILPINESTRILFIPDTFSICILRSQIIASKRIDSEFLSAVKEQLLT EANNIIEFVLDPEWA"

1. 140

/note="overlap with SC9920 Z46639 S. cerevisiae chromosome XIII cosmid 9920"

/complement(851..1312)

/gene="TIF11"

/complement(851..1312)

/gene="TIF11"

/note="YM8156.02c TIF11 gene, IF1A_YEAST, P3891, Initiation factor 1a"

/codon_start=1

/product="Tif1p"

/protein_id="CA89243.1"

/db_xref="GI:809083"

/db_xref="SWISS-PROT:P38912"

/translation="WGKNTKGKGRGKGRKNDSDGPKRELIYKEGQEVAYQITKMLGN GRVASCDFGKRMARHIGKRLKRYMGCODIILVSLDFDQDCDYVHKYLDIART LKNGELPENAKINETDNEGFESDDVNFEPGNADDEDEGEDELIDDI"

complement(1678..4842)

/gene="TPS3"

/complement(1678..4842)

/gene="TPS3"

/note="YM8156.03c, TPS3 gene, len: 1054, CAI: 0.21, SW:TPS3_YEAST P38426, alpha, alpha-crenolase-phosphate synthase and PIR:S27471, PS00213 Lipocalin signature; translation in TPS3_YEAST starts at amino acid 33 in this sequence"

/codon_start=1

/product="TPS3p"

/protein_id="CA89244.1"

/db_xref="GI:809084"

/db_xref="SWISS-PROT:P38426"

/translation="WTIIVASLFLPYTPQFADYVNSTAKLVESMIRKVDNNQELS NNKQSSSVTSASGSIYGLPOEAGINSEPCORANVGSPATGVNHNHNMESSEFOL FELNANATHANAGCIPANNPVSSGTSORPVEEFSPAPARVCSPOEASASISA SRSSAHNDLSSSLKMNPLNSFDSHPPRVRSKSAVITPVSKSPVDPAVDVAKY REEFQOQASLPMSKRVSGSTAGDSIASSSSLNRISQFQDNFIDTDSDEDIDSDLE TDATKRYNVPFGGYSNNAKLASLMSLRSYELFKHLPTIYDSKNGSLNNAVINIA AKTIVKEPVSVMGWTGIPITDELPHVECHKISKRLDQDFSSFPVVTDDITFGAYNVA

misc-feature

CDS

/note="PS00213 Lipocalin signature"

5199..6140

/note="YM8156.04, unknown, len: 313, CAI: 0.11, previously noted ORF, SW:YTP3_YEAST P38430"

/codon_start=1

/product="unknown"

/protein_id="CA89245.1"

/db_xref="GI:809085"

/db_xref="SWISS-PROT:P38430"

/translation="MKNLYDAHCYVITDPDTPFGDGGSGGTLACVASSNPFYDNNLT KILGRSTSKNDICVGFVHPVSHLSEFVSGRRKVSRYQVLYEKNEQSDSLVQVL PEPLDLEIKRENDLVSYIGIDKRLPLANGVYMNENKRLATYVAKLSHOE TVFRFWRLARHTSKPISIHVRCCHKANDICNEELIYHETIHLCHSYTGSEKELLG QMLKKPPDRIFVSLSKINKFDEEGALVRSIPSTILHETVDPIDNPDSYKAL TEOQLYNAQIARAMDETLDSQAALRYENFQFIK"

6392..6997

/note="YM8156.05, unknown, len: 301, CAI: 0.12, SW:YTP2_YEAST P38429 previously noted ORF"

/codon_start=1

/product="unknown"

/protein_id="CA89246.1"

/db_xref="GI:809086"

/db_xref="SWISS-PROT:P38429"

/translation="MARVYNNAETESGRPTGGGVASNNNGSCNNNGSNNNNNN NNNNNNSNNNNNGPTSSGRTNGKORITAAQOQYIKMLIETHIDNHPDLRPSHPM DEEYTAFLRKYKDHFDQVNDLQGLYGLSGSLAKTYSYKRNNGOGHDKRLHRR DIANVVRHHPDEHSIKETDCIPRIYKKNKKKFKMEFRG"

7278..7889

/note="YM8156.06, unknown, len: 301, CAI: 0.13, SW:YTP1_YEAST P38428, extends previously noted ORF"

/codon_start=1

/product="unknown"

/protein_id="CA89247.1"

/db_xref="GI:809087"

/db_xref="SWISS-PROT:P38428"

/translation="MEDSRLLITLILYGVYFLKFFQSNQHPNSAQRISANGVNAHGR POGSTONALRRGTGVNGHPYTTQMVQVLAFLRHEQRIYSLENTGSVEYETERY LRGEFEPPEPGEFSPAPMGMAVNDNNAAGGEPNDRKKNMCAENLIDKFEHVDN EDMSNLSFKDIDIERKRLYQARKNLETLQSDKDLQSLIT"

complement(8013..9398)

/note="YM8156.07c, unknown, len: 461, CAI: 0.11"

/codon_start=1

/product="unknown"

/protein_id="CA89248.1"

/db_xref="GI:809088"

/db_xref="SWISS-PROT:Q03508"

/translation="MEEFEEFRKRGKSGNHRVLRKNSACCELAAPVEPEPHAI TKLIHYDNDNTLFPATPGTEBOLYRELLNLTSLTPNGMNEGFLQAAIEISKTK PRYSNNAADIVKLAESYSKADLTISYIYEGEERFHLIHAQOTASHNKCSNER FENAVYCKARRAIEYTSYKKEKRLDLEIYPSLRELSITDRHQIDAFSFFSISL LPRKMSAIVPPEPTKSLPREQELMVDVNRKNNSQLSTVQFDRRTPRQIDYIL CTAGHRLLSIEVYIKYLRKRCRFRRLYHPYIIPCAEFGKIDIPALEIKAVMSND TRTDSKRYKOHISQIYRLBOPGACIYHFOVYTLAVIYSAHNNRRKPLEYEFKAPR NRYETFEPEYVGHYKKNRIDLEVEYERLINCKEDIMWPLDNTIPIKAFGRF AKLAIFCSNR"

9991..12852

/note="YM8156.08, unknown, len: 953, CAI: 0.20, probable integral membrane glycoprotein"

/codon_start=1

/product="unknown"

/protein_id="CA89249.1"

/db_xref="GI:809089"

/db_xref="SWISS-PROT:Q03509"

/translation="MEEFEEFRKRGKSGNHRVLRKNSACCELAAPVEPEPHAI TKLIHYDNDNTLFPATPGTEBOLYRELLNLTSLTPNGMNEGFLQAAIEISKTK PRYSNNAADIVKLAESYSKADLTISYIYEGEERFHLIHAQOTASHNKCSNER FENAVYCKARRAIEYTSYKKEKRLDLEIYPSLRELSITDRHQIDAFSFFSISL LPRKMSAIVPPEPTKSLPREQELMVDVNRKNNSQLSTVQFDRRTPRQIDYIL CTAGHRLLSIEVYIKYLRKRCRFRRLYHPYIIPCAEFGKIDIPALEIKAVMSND TRTDSKRYKOHISQIYRLBOPGACIYHFOVYTLAVIYSAHNNRRKPLEYEFKAPR NRYETFEPEYVGHYKKNRIDLEVEYERLINCKEDIMWPLDNTIPIKAFGRF AKLAIFCSNR"

9991..12852

/note="YM8156.08, unknown, len: 953, CAI: 0.20, probable integral membrane glycoprotein"

/codon_start=1

/product="unknown"

/protein_id="CA89250.1"

/db_xref="GI:809090"

/db_xref="SWISS-PROT:Q03510"

/translation="MEEFEEFRKRGKSGNHRVLRKNSACCELAAPVEPEPHAI TKLIHYDNDNTLFPATPGTEBOLYRELLNLTSLTPNGMNEGFLQAAIEISKTK PRYSNNAADIVKLAESYSKADLTISYIYEGEERFHLIHAQOTASHNKCSNER FENAVYCKARRAIEYTSYKKEKRLDLEIYPSLRELSITDRHQIDAFSFFSISL LPRKMSAIVPPEPTKSLPREQELMVDVNRKNNSQLSTVQFDRRTPRQIDYIL CTAGHRLLSIEVYIKYLRKRCRFRRLYHPYIIPCAEFGKIDIPALEIKAVMSND TRTDSKRYKOHISQIYRLBOPGACIYHFOVYTLAVIYSAHNNRRKPLEYEFKAPR NRYETFEPEYVGHYKKNRIDLEVEYERLINCKEDIMWPLDNTIPIKAFGRF AKLAIFCSNR"

9991..12852

/note="YM8156.08, unknown, len: 953, CAI: 0.20, probable integral membrane glycoprotein"

/codon_start=1

/product="unknown"

/protein_id="CA89251.1"

/db_xref="GI:809091"

/db_xref="SWISS-PROT:Q03511"

/translation="MEEFEEFRKRGKSGNHRVLRKNSACCELAAPVEPEPHAI TKLIHYDNDNTLFPATPGTEBOLYRELLNLTSLTPNGMNEGFLQAAIEISKTK PRYSNNAADIVKLAESYSKADLTISYIYEGEERFHLIHAQOTASHNKCSNER FENAVYCKARRAIEYTSYKKEKRLDLEIYPSLRELSITDRHQIDAFSFFSISL LPRKMSAIVPPEPTKSLPREQELMVDVNRKNNSQLSTVQFDRRTPRQIDYIL CTAGHRLLSIEVYIKYLRKRCRFRRLYHPYIIPCAEFGKIDIPALEIKAVMSND TRTDSKRYKOHISQIYRLBOPGACIYHFOVYTLAVIYSAHNNRRKPLEYEFKAPR NRYETFEPEYVGHYKKNRIDLEVEYERLINCKEDIMWPLDNTIPIKAFGRF AKLAIFCSNR"

9991..12852

/note="YM8156.08, unknown, len: 953, CAI: 0.20, probable integral membrane glycoprotein"

/codon_start=1

```

/note="M8156.09, IPP2 gene (PPA), len: 310, CAT: 0.14,
SW:1Y2.YEAST, P28239, Inorganic pyrophosphatase,
mitochondrial; contains P500387 Inorganic pyrophosphatase
signature"
/codon_start=1
/product="ipp2p"
/protein_id="CA89250.1"
/db_xref="GI:809090"
/db_xref="SWISS-PROT:P28239"
/transtation="NNLEPMNLTKASRIERLKQTLNLTISNNHROFTIOGSKTT
GFKRYLLLNQEVSSFDVLDLNEHKRTVMIVIEPWTIGKEISLELFPNPV

```

| | | | |
|----|-------|--|-------|
| OY | 1 | tcgaactgttgatgaaacaaagaaactctcaaacgactcgaagcttggtaagatgatttcca | 60 |
| Db | 27785 | TGAAGTGTGATGAAACAAAGAAAGAAAGCTGACAAAGCATCAAAAGCTTGTGAGATGATTTCCA | 27726 |
| OY | 61 | ttcaaaaattctgtggaata tgaaacccaagt tccaacaagaattcttcctgaaataa tgaagtt | 120 |
| Db | 27725 | TTCAAAATTTGTGAATATGAAACCAAGTTCACAAAGATTTCTTGATTAATAGAGTT | 27666 |
| OY | 121 | ataataatlgaaatgggggaagacgctgaagccaaaggaactgaaatttacaagatttattac | 180 |
| Db | 27665 | ATATAATVTGAAATAGGAGAGACCGTAGGCGCAAAAGAACTGAATTTATACGATTTATATAC | 27606 |
| OY | 181 | ttcacagctgtgttaacgatatcgaagaacttgcgaacttgaagttgaagttgttcaaaact | 240 |
| Db | 27605 | TTCAACAGCTGTGTTAACGATATGAAAGCATTTGCAACTTGAAAGTTAGCTTCACAACTT | 27546 |
| OY | 241 | atccaaacatctttaaaaaaagcgtcagtaaccttaaaaaacaactcaaaagacgcattac | 300 |
| Db | 27545 | ATCCAAACATTTTTTAAAAAACAAGCTGATGATACCTTAATAACACACCTCAAAACACGCAATTATAC | 27486 |
| OY | 301 | ctgaagattccaatatccgacacagcgtgaacggtatgattcatatcgtccgcagctgaagctcatt | 360 |
| Db | 27485 | TGAGGATTCATATTCGCACACAGGTAAAGCGTAGTTCATGTGTGCCCAAGTTCAGGCTCATTT | 27426 |
| OY | 361 | aacgcacgtaactcttcataccaagaagttattattttaccatacgcgtgtagctgtctac | 420 |
| Db | 27425 | AACGCCATTAACCTTTTCATCCAAAGATGATTTATTTTACCTAGAGGAGTGCCTGCTAC | 27366 |
| OY | 421 | ttccctggaatttacaagccagatgcttcataaaagggttgaagttgctcccttacaagta | 480 |
| Db | 27365 | TTCCCTGGAATTTAAGACCAAGATGTTTCATATAAAGGGTTAGGATTTGCTCCTTTACAGTA | 27306 |

| | | | |
|----|-------|--|-------|
| OY | 481 | caacgagacatlaatgfygaactfygaattttaaagaacatlaaggaacattcaatcc | 540 |
| Db | 27305 | CAAAAGAACATTAATGGAACCTTGAAATTATATAGACATTAGAAACTTAATTC | 27246 |
| OY | 541 | aagtttgaagcbycccaigtctgfyaggttttatctcgcctcgaagtaagatlaattga | 600 |
| Db | 27245 | AAGTTTGAAGCTCCATGTTAGTGTATTTATTCGCTTGAGTAATGATTAAATTGGA | 27166 |
| OY | 601 | aaacacatctggcgtagcgaagatgtatatacccatcttcglttagcgaagycacaata | 660 |
| Db | 27185 | AAACCATCTTGGCGTAGCGAAGATGAAATGCCCTTTCTGTTAGCAAGTACAAAAATA | 27126 |
| OY | 721 | gcttaaatgaagatlaatacgctcccttgtaatacccatlaataaattcctt | 780 |
| Db | 27065 | GCTTATATATGAAGTATTTATTCGGCTCTTGATTCGCCATTTATATAAATTCCTT | 27066 |
| OY | 781 | taaatgcatcttcgtygtcccttgltgctcctgtaatttttttttggaccactg | 840 |
| Db | 27005 | TAAATGCATTTTCGCGCTCTTTGTTGCTGCTGTAATTTTTTTGGACCACWG | 26946 |
| OY | 841 | gatggaacacttfgatgatttttacccttatattgaattacaaataatcgagatt | 900 |
| Db | 26945 | GATGGAACACCTTGATATTTTATTAACCTTATTTAAGTAACTAAATATCGAGATT | 26886 |
| OY | 901 | caggaacaaacatagaattctcttcgccaagaaataaaccgaataaattgagct | 960 |
| Db | 26885 | CAGGACAAACATGGAATTTCTTTGTCAAGAAAAATAAACGAATTAATGATGCTT | 26826 |
| OY | 961 | tgaactcgaactgtctgcatagagagaacgaacgaacaaigtacaaattctcaaat | 1020 |
| Db | 26825 | TGACTACGACTGCTGTGATAGAGAGAACGACGACCAATGCTAACATTCCTCAACTT | 26766 |
| OY | 1021 | tctagtgctctatacaatctccatatcctatgctacgcgtatatccgaacttcaaac | 1080 |
| Db | 26765 | TCTATGCTCTTATACATCTCCATATCCTATGTGCTACCGCATATGGAACCTCCAAAC | 26706 |
| OY | 1081 | agaagggctaaacagagataatcccgaaagatlaaactctgcagcaaaaaacttaacat | 1140 |
| Db | 26705 | AGAAAGGCTTAACAGAGATAATCCCGAAGATTAAATCTGCCATGCCAAAACTTACAT | 26646 |
| OY | 1141 | fatgctaaattccaaaccttttttgtygccttttttacaacttccaattcctagtaaac | 1200 |
| Db | 26645 | TATGCTAATTTCCAAACCTTTTGTGCTCTTTTACAACTCAATATCTAGTAACAC | 26586 |
| OY | 1201 | ttcaacataaagttccaagagacatlttttaggccttaggtattatccgaagtataagc | 1260 |
| Db | 26585 | TTCACATATTAAGTTTCAAGAGCGCACTTTTATAGGCTATAGTATTAATCCACAGGTATTAAGC | 26526 |
| OY | 1261 | tgcacttgcacaaccttggcaatcgaagccagttctgtgaagaacttaacgaaaatgtytgc | 1320 |
| Db | 26525 | TGCATTGCCAAACCTTGGCAATTGAGCCAGTGTGTGAAGACTTAACGAATATGCTTGC | 26466 |
| OY | 1321 | gatgttattgacctatactgtggaacccgttttagattttgtatataatcaatttaaa | 1380 |
| Db | 26465 | GATGTTATTGAACCTTATATGTGGAGCCGTTTATAGATTTTGATATATATCAATTAATAA | 26406 |
| OY | 1381 | ttcaaaagagcctataacttgaagaatttttacaagaatctctgaaatattggagttcaag | 1440 |
| Db | 26405 | TTCAAAAGAGCTTATACCTTGAAGATTTTACCATTAATCTCGAATTAATTTGGAGTTTGA | 26346 |
| OY | 1441 | gaatttatatttgcacaataaactgaggaagaatactttaacgcgtcaatgctttgacac | 1500 |
| Db | 26345 | GAAATTTTATTTGCACCAATTAACCTGAGGAATATTTTACACGTCATATGCTTTTAACTAC | 26286 |
| OY | 1501 | gtacttlaacctlaaccgcatctcgcaactaagcatatacaaaagttatttttggcaaccatc | 1560 |
| Db | 26285 | GTACTTAACCTTAATACCGCAATTCGCACATTAAGCTATCAACAGTTATTTTGGCAACCATC | 26226 |
| OY | 1561 | gcttttttttggacttgcgcgcgcacacacatgcttatgagcaattcaaggaagctccat | 1620 |

[illegible]

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 12, 2002, 09:26:44 ; Search time 17.02 Seconds

(without alignments)
975.862 Million cell updates/sec

Title: US-09-165-460a-2

Perfect score: 2342
Sequence: 1 MEDLKTILDHPNIPWKLIS.....HPTLAERSTALDYSEKKKN 453

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : SwissProt_39.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|-------------|--------------------|
| 1 | 2336 | 99.7 | 453 | ST24_YEAST | P47154 saccharomyc |
| 2 | 903 | 38.6 | 474 | ST24_SCHPO | Q10071 schizosacch |
| 3 | 698.5 | 29.8 | 475 | FAC1_HUMAN | O75844 homo sapien |
| 4 | 328 | 14.0 | 426 | YHFN_BACSU | P40769 bacillus su |
| 5 | 161 | 6.9 | 286 | HTPX_PASMU | P57846 pasteurella |
| 6 | 150.5 | 6.4 | 310 | HTPX_HELPJ | O82k44 heliobacte |
| 7 | 150 | 6.4 | 335 | HTPX_ARCFU | O30004 archaeoglob |
| 8 | 147.5 | 6.3 | 283 | HTPX_HAEIN | P44840 haemophilus |
| 9 | 142.5 | 6.1 | 310 | HTPX_HELPY | O25582 heliobacte |
| 10 | 139 | 5.9 | 366 | HTPX_AERPE | O25582 aeropyrum p |
| 11 | 138.5 | 5.9 | 258 | HTPX_METHH | O58669 methanobact |
| 12 | 130.5 | 5.6 | 292 | HTPX_PYRHO | O58997 pyrococcus |
| 13 | 130.5 | 5.6 | 317 | HTPX_THERAC | O2Hj22 thermoplasm |
| 14 | 127.5 | 5.4 | 2136 | YCF2_MARPO | P09975 marichantia |
| 15 | 120 | 5.1 | 292 | HTPX_BUCAI | P57406 buchiera ap |
| 16 | 119.5 | 5.1 | 2294 | YCF2_ARATH | P67676 arabidopsis |
| 17 | 119 | 5.1 | 297 | HTPX_STRGC | O30795 streptococc |
| 18 | 116 | 5.0 | 293 | HTPX_ECOLI | P33894 escherichia |
| 19 | 113 | 4.8 | 284 | HTPX_MERJA | O59076 methanococc |
| 20 | 113 | 4.8 | 298 | HTPX_BACSU | O31657 bacillus su |
| 21 | 109 | 4.7 | 506 | YCX2_EUDGR | P31921 euglena gra |
| 22 | 109 | 4.7 | 595 | YCB6_MERJA | O58682 methanococc |
| 23 | 107.5 | 4.6 | 1032 | MT18_YEAST | P40469 saccharomyc |
| 24 | 106.5 | 4.5 | 646 | HTPX_HAREV | O37319 heliothis a |
| 25 | 106.5 | 4.5 | 1029 | YFC5_YEAST | P43571 saccharomyc |
| 26 | 105.5 | 4.5 | 873 | COX1_ACACA | O37370 acanthamoeb |
| 27 | 105 | 4.5 | 1427 | SRB8_YEAST | P25648 saccharomyc |
| 28 | 104.5 | 4.5 | 585 | YHDB_YEAST | P58732 saccharomyc |
| 29 | 104.5 | 4.5 | 1100 | LOS1_YEAST | P33478 saccharomyc |
| 30 | 104 | 4.4 | 302 | HTPX_AQUAE | O67798 aquifex aeo |
| 31 | 104 | 4.4 | 530 | CP14_CHICK | P79760 gallus gall |
| 32 | 103.5 | 4.4 | 185 | GU01_RAT | P35894 rattus norv |
| 33 | 103.5 | 4.4 | 450 | VATH_SCHPO | O14265 schizosacch |

ALIGNMENTS

| RESULT | 1 | ST24_YEAST | STANDARD | PRT | 453 AA |
|--------|---|------------|----------|-----|--------|
| AC | P47154: | | | | |
| DT | 01-FEB-1996 (Rel. 33, Created) | | | | |
| DT | 01-FEB-1996 (Rel. 33, Last sequence update) | | | | |
| DT | 20-AUG-2001 (Rel. 40, Last annotation update) | | | | |
| DE | CAAX PRENYL PROTEASE 1 (EC 3.4.24.-) (PRENYL PROTEIN-SPECIFIC DE ENDOPEPTASE 1) (PSEPT 1) (A-FACTOR CONVERTING ENZYME). | | | | |
| GN | STE24 OR APC1 OR YTR117W OR J2032. | | | | |
| OS | Saccharomyces cerevisiae (Baker's yeast). | | | | |
| OC | Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; | | | | |
| OC | Saccharomycetales; Saccharomycetaceae; Saccharomycetes. | | | | |
| OX | NCBI_TaxID=4932; | | | | |
| RP | SEQUENCE FROM N.A. | | | | |
| RA | Rose M., Koetter P., Entian K.D.; | | | | |
| RL | Submitted (SEP-1995) to the EMBL/GenBank/DBJ databases. | | | | |
| RN | [2] | | | | |
| RP | SEQUENCE FROM N.A. AND FUNCTION. | | | | |
| RA | MEDLINE=97167681; PubMed=9015299; | | | | |
| RT | Fujimura-Kamada K., Nouvet F.J., Michaelis S.; | | | | |
| RT | "A novel membrane-associated metalloprotease, Ste24p, is required for the first step of NH2-terminal processing of the yeast a-factor precursor." | | | | |
| RT | J. Cell Biol. 136:271-285(1997). | | | | |
| RN | [3] | | | | |
| RP | FUNCTION. | | | | |
| RA | MEDLINE=97218305; PubMed=9065405; | | | | |
| RT | Boyarchuk V.L., Ashby M.N., Rine J.; | | | | |
| RT | "Modulation of Ras and a-factor function by carboxyl-terminal proteolysis." | | | | |
| RL | Science 275:1796-1800(1997). | | | | |
| RN | [4] | | | | |
| RP | FUNCTION. | | | | |
| RA | MEDLINE=98365461; PubMed=9700155; | | | | |
| RT | Tam A., Nouvet F.J., Fujimura-Kamada K., Slunt H., Sisodia S.S., Michaelis S.; | | | | |
| RT | "Dual roles for Ste24p in yeast a-factor maturation: NH2-terminal proteolysis and COOH-terminal CAAX processing." | | | | |
| RL | J. Cell Biol. 142:635-649(1998). | | | | |
| RN | [5] | | | | |
| RP | FUNCTION. | | | | |
| RA | MEDLINE=98393572; PubMed=9725832; | | | | |
| RT | Boyarchuk V.L., Rine J.; | | | | |
| RT | "Roles of prenyl protein proteases in maturation of Saccharomyces cerevisiae a-factor." | | | | |
| RL | Genetics 150:95-101(1998). | | | | |
| RN | [6] | | | | |
| RP | FUNCTION. | | | | |
| RA | MEDLINE=20158928; PubMed=10692417; | | | | |
| RT | Schmidt W.K., Tam A., Michaelis S.; | | | | |
| RT | "reconstitution of the Ste24p-dependent N-terminal proteolytic step in yeast a-factor biogenesis." | | | | |
| RL | J. Biol. Chem. 275:6227-6233(2000). | | | | |
| RN | [7] | | | | |

| | | | | | | |
|----|-------|-----|------|---|------------|--------------------|
| 34 | 103 | 4.4 | 645 | 1 | VP74_NPYAC | P15663 autographa |
| 35 | 103 | 4.4 | 1101 | 1 | KOE5_YEAST | O08217 saccharomyc |
| 36 | 102 | 4.4 | 577 | 1 | TH71_SCHPO | O09684 schizosacch |
| 37 | 102 | 4.4 | 2493 | 1 | YBA4_YEAST | P35194 saccharomyc |
| 38 | 100.5 | 4.3 | 260 | 1 | YMI6_RECAM | O21266 reclinomona |
| 39 | 100.5 | 4.3 | 1040 | 1 | RIK1_SCHPO | O10426 schizosacch |
| 40 | 100 | 4.3 | 1325 | 1 | MRP4_HUMAN | O15439 homo sapien |
| 41 | 99.5 | 4.2 | 309 | 1 | O7AA_HUMAN | O76100 homo sapien |
| 42 | 99 | 4.2 | 485 | 1 | YCI1_KLEPN | O48457 klebsiella |
| 43 | 99 | 4.2 | 1024 | 1 | RPOB_PLAFA | P21421 plasmodium |
| 44 | 98.5 | 4.2 | 313 | 1 | NU2M_RHTSA | O99817 thripcephal |
| 45 | 98.5 | 4.2 | 417 | 1 | O65A_DROME | P82982 drosophila |

RE CHARACTERIZATION.
 RA MEDLINE-20285442; PubMed-10825201;
 RA Treiblood C.E., Boyartchuk V.L., Picooglou E.A., Rozema D.,
 RA Poulter C.D., Rhine J.;
 RT "The CAAX proteases, Atc1p and Rce1p, have overlapping but distinct
 RT substrate specificities." Mol. Cell. Biol. 20:4381-4392(2000).
 RL (8)
 RN SUBCELLULAR LOCATION.
 RP MEDLINE-98409630; PubMed-9736709;
 RA Schmidt W.K., Tam A., Fujimura-Kameda K., Michaelis S.;
 RT "Endoplasmic reticulum membrane localization of Rce1p and Ste24p,
 RT yeast proteases involved in carboxyl-terminal CAAX protein processing
 RT and amino-terminal A-factor cleavage." J.
 RL Proc. Natl. Acad. Sci. U.S.A. 95:11175-11180(1998).
 CC -1- FUNCTION: PROTEOLYTICALLY REMOVES THE C-TERMINAL THREE RESIDUES OF
 CC FARNESYLATED A-FACTOR MATING PHEROMONE. ALSO ACTS TO CLEAVE THE N-
 CC TERMINAL EXTENSION OF THE PHEROMONE. DOES NOT ACT ON RAS.
 CC -1- COFACTOR: BINDS ONE ZINC ION (PROBABLE).
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. ENDOPLASMIC
 CC RETICULUM.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M48 (ZINC
 CC METALLOPROTEASE).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: 249617; CAAB9647.1; -
 DR EMBL: U77137; AAB38271.1; -
 DR MEROPS: M48.001; -
 DR SCD: S0003878; STE24.
 DR InterPro: IPR001915; Peptidase_M48.
 DR InterPro: IPR000130; Zn_MTpeptidase.
 DR Pfam: PF01435; Peptidase_M48; 1.
 DR PROSITE: PS00142; ZINC_PROTEASE; 1.
 KW Hydrolyase; Metalloprotease; Zinc; Transmembrane;
 KW Endoplasmic reticulum; Pheromone response.
 FT TRANSMEM 13 33 POTENTIAL.
 FT TRANSMEM 90 110 POTENTIAL.
 FT TRANSMEM 122 142 POTENTIAL.
 FT TRANSMEM 168 188 POTENTIAL.
 FT TRANSMEM 198 218 POTENTIAL.
 FT TRANSMEM 307 327 POTENTIAL.
 FT TRANSMEM 358 378 POTENTIAL.
 FT METAL 297 297 ZINC (CATALYTIC) (BY SIMILARITY).
 FT ACT_SITE 298 298 ZINC (CATALYTIC) (BY SIMILARITY).
 FT METAL 301 301 ZINC (CATALYTIC) (BY SIMILARITY).
 FT METAL 390 390 ZINC (CATALYTIC) (BY SIMILARITY).
 FT ACT_SITE 394 394 PROTON DONOR (BY SIMILARITY).
 SO SEQUENCE 453 AA; 52324 MW; 331C9AE2D7C99DA CRC64;

Query Match 99.7%; Score 2336; DB 1; Length 453;
 Best Local Similarity 99.8%; Pred. No. 3.7e-160;
 Matches 452; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MFDTKTLIDHNPIMKLIISGFSIAOFSESYLYROYOKLSETRKLPVLEDEIDDETFH 60
 DB 1 MEDKTLIDHNPIMKLIISGFSIAOFSESYLYROYOKLSETRKLPVLEDEIDDETFH 60
 QY 61 KSRNYSRAKAFSTFGDYNNLAQKLVFKYDLFPKIMHVAVSLNAVLPVRFHNVSTVAQ 120
 DB 61 KSRNYSRAKAFSTFGDYNNLAQKLVFKYDLFPKIMHVAVSLNAVLPVRFHNVSTVAQ 120
 QY 121 SLCEGLLSLSTVDLPVLSYSHVLEERGFNKLTPQVLTMTMKSILTAIYAGRIIL 180
 DB 121 SLCEGLLSLSTVDLPVLSYSHVLEERGFNKLTPQVLTMTMKSILTAIYAGRIIL 180

QY 181 YLFKIFDKFPFDLMYIMVLEFVQVLIAMTIIPIFMPEMFKPTLEDGLKKSIESLA 240
 DB 181 YLFKIFDKFPFDLMYIMVLEFVQVLIAMTIIPIFMPEMFKPTLEDGLKKSIESLA 240
 QY 241 DRVGFPDLKIVIDGSKSSSSNAFTGLPPTSKRIYVFDLVNSNSTDEITAVLAHEIG 300
 DB 241 DRVGFPDLKIVIDGSKSSSSNAFTGLPPTSKRIYVFDLVNSNSTDEITAVLAHEIG 300
 QY 301 HMOKNHIYNNVIFSOHLFIETSLFTSYRNTSPFNFGFLEKSTGSEFVDPVITKEPPI 360
 DB 301 HMOKNHIYNNVIFSOHLFIETSLFTSYRNTSPFNFGFLEKSTGSEFVDPVITKEPPI 360
 QY 361 IIGFPLFNDLTPLECAMQFVMSLISRTHEYOADAVAKKLGKONLCRALDLOIKNLST 420
 DB 361 IIGFPLFNDLTPLECAMQFVMSLISRTHEYOADAVAKKLGKONLCRALDLOIKNLST 420
 QY 421 MNVDPLVSSYHSHTLAERSTALDYVSEKKRN 453
 DB 421 MNVDPLVSSYHSHTLAERSTALDYVSEKKRN 453
 RESULT 2
 ST24_SCHPO STANDARD: PRT: 474 AA.
 ID ST24_SCHPO
 AC Q10071;
 DT 01-FEB-1996 (Rel. 33, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE PROBABLE CAAX PRENILE PROTEASE 1 (EC 3.4.24.-) (PRENILE PROTEIN-
 DE SPECIFIC ENDOPEPTIDASE 1) (PSPSP 1).
 GN SPAC3H1.05.
 OS Schizosaccharomyces pombe (fission yeast).
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;
 OC Schizosaccharomyces.
 OX NCBI_TaxID=4896;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=972;
 RA Gentiles S., Churcher C.M., Barrell B.G., Rajandream M.A., Walsh S.V.;
 RL Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: PROTEOLYTICALLY REMOVES THE C-TERMINAL THREE RESIDUES OF
 CC FARNESYLATED PROTEINS (BY SIMILARITY).
 CC -1- COFACTOR: BINDS ONE ZINC ION (POTENTIAL).
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. ENDOPLASMIC
 CC RETICULUM (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M48 (ZINC
 CC METALLOPROTEASE).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: Z68144; CA92258.1; -
 DR MEROPS: M48.001; -
 DR InterPro: IPR001915; Peptidase_M48.
 DR InterPro: IPR000130; Zn_MTpeptidase.
 DR Pfam: PF01435; Peptidase_M48; 1.
 DR PROSITE: PS00142; ZINC_PROTEASE; 1.
 KW Hypothetical protein; Hydrolase; Metalloprotease; Zinc; Transmembrane;
 KW Endoplasmic reticulum.
 FT TRANSMEM 103 123 POTENTIAL.
 FT TRANSMEM 196 216 POTENTIAL.
 FT TRANSMEM 230 250 POTENTIAL.
 FT TRANSMEM 344 364 POTENTIAL.
 FT TRANSMEM 381 401 POTENTIAL.
 FT METAL 332 332 ZINC (CATALYTIC) (BY SIMILARITY).
 FT METAL 333 333 ZINC (CATALYTIC) (BY SIMILARITY).
 FT ACT_SITE 336 336 ZINC (CATALYTIC) (BY SIMILARITY).
 FT METAL 336 336


```

Db 141 WSLYNTFVIEERHGFNOQTLGFMKDAIKKFVVTQCLLPVSSLLIYIKIGDYFIYA 200
Qy 199 MPFLPVQOLAMTIIIPVFMPMFNKFTPLDEGEIKSSTSLADRCFPLDKIFVIDGSKR 258
Db 201 WFLTVLVSALVLTADYADLPFDKFTPLPEGLKEELVEMAKSIDPLTKVYVEGSKR 260
Qy 259 SSHSNAYFTGLPFTSKRIYLFDTLVNSNS----- 287
Db 261 SSHSNAYFTGLF-FKNKRIYLFDTLVNSNS----- 319
Qy 288 -----TDEITAVLAHEIGHQWOKNHNIVNVIFSQLHTFLIFSLTSIYRNTSEFYNTGFF 341
Db 320 KRQGCNNEEVLAVLGHGELHMKLGHYVKNIIISQMSFICFELFAVLIGRKELEFAFGFY 379
Qy 342 LKRSKTSFDPVITKEFPIIIGFM-LFNDLLTPLECAMQFVMSLSRTHYQADAVAKKL 400
Db 380 DSO-----PFLIGLIIIFQIFSPYNEVLSFCLTVLSRFEQADAFAKKL 425
Qy 401 GYKONICRALIDQIKNLSTMNVDPLYSYHYSHPTLAERSTAL 444
Db 426 GKAKDLXSLIKLNKNDLGFVSDMLFMSMHSHPLERLQAL 469

RESULT 4
YHFN_BACSU STANDARD: PRT: 426 AA.
ID PA0769: 007612:
AC 01-FEB-1995 (Rel. 31, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DE 20-AUG-2001 (Rel. 40, Last annotation update)
DN PUTATIVE METALLOPROTEASE YHFN (EC 3.4.24.-) (PSP23).
GN YHFN.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RA Noback M.A., Terpstra P., Holsappel S., Venema G., Bron S.;
RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 1-49 FROM N.A.
RX MEDLINE=89108019; PubMed=3145906;
RT Smith H., de Jong A., Bron S., Venema G.;
RT "Characterization of signal-sequence-coding regions selected from the
RL Bacillus subtilis chromosome.";
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M48 (ZINC
CC METALLOPROTEASE).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL: Y14083; CAA74535.1;
CC EMBL: M22914; AAM22830.1;
CC EMBL: 299109; CABI2869.1;
CC Subtilist: BG11029; yhfn.
CC InterPro: IPR001915; Peptidase_M48.
CC InterPro: IPR000130; Zn_MTPepdse.
CC Pfam: PF01435; Peptidase_M48; 1.
CC PROSITE: PS00142; ZINC_PROTEASE; 1.
CC Hypothetical protein: Hydrolase; Metalloprotease; Zinc;
CC Complete proteome.
CC METAL 277 277 ZINC (CATALYTIC) (POTENTIAL).
CC ACT_SITE 278 278 POTENTIAL.
CC METAL 281 281 ZINC (CATALYTIC) (POTENTIAL).

```

```

FT CONFLICT 43 49 ASFMKPS -> LESTAOA (IN REF. 2).
SQ SEQUENCE 426 AA; 48874 MW; 0F69E7EFAA80143 CRC64.

Query Match 14.0%; Score 328; DB 1; Length 426;
Best local similarity 26.9%; Pred. No. 1.7e-16;
Matches 108; Conservative 76; Mismatches 152; Indels 66; Gaps 14;

Qy 81 LAQKLVFIKYDF-----PKIMMAVSL-----NAVLPVRHNMSTVAQSICLF 124
Db 53 VAEQYSNVKNFLFPIGVPLDMFLFVLLVSGYSKKIKKIKWEAAVPRF-----LQTVGF 106
Qy 125 LGLLSLSLTVLPLPSYSHFELEKFGFNKLTQVLTMDTKSTLAVYAG--PILY 181
Db 107 VFLSLITLTVLPDWMIGYQVSLD-VNISTOTTSMAKDOVISWISFIFLCLVLFY 165
Qy 182 LFLKIDKFPPELWYINW---FLFVQOLAMTIIIPVFMFMFNKFTPLDEGEIKSIE 237
Db 166 WLIRHKKKMWLYAWLTVPFSLFLEFIQ-----PYIDPLVNDPFLKKELESKIL 218
Qy 238 SLADRVGFPDLKIFVIDGSKRSNNAFTGLPFTSKRIYLFDTLVNSNSTDEITAVLAH 297
Db 219 ELADANIPADHVEVNNSEKTNALNVAVTGIG-ANKRIYIMDTTLNKLDSSELLIFIGH 277
Qy 298 EIGHQWOKNHNIVNVIFSQLHTFLIFSL-----FTSIYRNTSEFYNTGFFLEKSTGSPVDP 352
Db 278 EMGHVYMKH-----YIGLAGLLVSLAGFYIDKLYKRYATRLTSMHLEGRHDLALP 332
Qy 353 VITKEFPIIIGFMFNNDLTPLECAMQFVMSLSRTHYQADAVAKKL-GYKONICRALI 411
Db 333 L-----LFLFSVLSFAVTPEFSNA-----VSRQGEKKAKQYGIETLENBAVKTFO 379
Qy 412 DQIKNLSTMNVDPLYSYHYSHPTLAERSTALDVSSEKKN 453
Db 380 DLAVGLSQVDPVPLVVKIFRGSHPSIMERIOH---AKREEN 417

RESULT 5
HHPX_PASMU STANDARD: PRT: 286 AA.
ID HTPX_PASMU
AC P57846;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DE 20-AUG-2001 (Rel. 40, Last annotation update)
DN PROBABLE PROTEASE HTPX HOMOLOG (EC 3.4.24.-).
GN HTPX OR PMO468.
OS Pasteurella multocida.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Pasteurella.
OX NCBI_TaxID=747;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PM70;
RX MEDLINE=21145866; PubMed=11248100;
RA May B.J., Zhang Q., Li L.L., Pautsian M.L., Whittam T.S., Kapur V.;
RT "Complete genomic sequence of Pasteurella multocida pm70.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465(2001).
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M48 (ZINC
CC METALLOPROTEASE).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL: AE006083; AAK02552.1;
CC InterPro: IPR001915; Peptidase_M48.
CC InterPro: IPR000130; Zn_MTPepdse.
CC Pfam: PF01435; Peptidase_M48; 1.

```


DR PROSITE: PS00142; ZINC_PROTEASE; FALSE_NEG.
 KW Transmembrane: Hydrolase; Metalloprotease; zinc; Complete proteome.
 FT TRANSMEM 4 24 POTENTIAL.
 FT TRANSMEM 33 53 POTENTIAL.
 FT TRANSMEM 186 206 POTENTIAL.
 FT METAL 139 139 ZINC (CATALYTIC) (BY SIMILARITY).
 FT ACT_SITE 140 140 BY SIMILARITY.
 FT METAL 143 143 ZINC (CATALYTIC) (BY SIMILARITY).
 SQ SEQUENCE 286 AA; 31344 MW; 58FAA7841C300BCF CRC64;

Query Match 6.9%; Score 161; DB 1; Length 286;
 Best Local Similarity 24.8%; Pred. No. 9.4e-05;
 Matches 76; Conservative 53; Mismatches 128; Indels 50; Gaps 13;

DB 15 LVFENIIILSTGICQADATGLMALALFGSGS-----LISLFSLKTMALFVAGAEVIK 68
 OY 160 LMITMISLTLAVAGGILYLFKIDKPTDFLWYIMVFLVYQILAMTIIPVIMP 219
 DB 220 MENKFTPLEDGE--LKRSIESLADRVGFLDKIFVIDGSKSSHSNAVETGIPFSKRIV 277
 DB 69 Q-----PRNDMERLVNTVRSOAEKALPMDVAIV---HSEDVNAEATG-PSKNSSLV 118
 OY 278 LFDT-LVNSNSTDEITAVLAHEIGHQKNHIVNYIF-SQHTFLIFSLFSTISYNTSEY 335
 DB 119 AVSTGLLAKMTQDEAEVLAHEVAHAIKNGDMVTMLLGVLNTEFVIF--VSRMIKAVYS 176
 OY 336 NTEGFFLEKSTGSEVDVYITKEPFIILGFMLENDLPLECAMQFVMSLISFHEYQADA 395
 DB 177 NNDG---ESSSTGIYF--LVSWLEILFGFL-----ASMIAMFESRYREFRADA 219
 OY 396 YAKKIGYONLCRALIDLOIKN-----ISTMNVDPL---YSSYHSHPLARSTAL 444
 DB 220 GSAKLVGHKMAIALQRLQRLHEPQELGQALFAINCKRGQALALFMSHPLEKRIAL 279
 OY 445 DYVSEKK 451
 DB 280 QQLDSFK 286

RESULT 6
 HTPX_HELPJ STANDARD; PRT; 310 AA.
 AC 09KKS4:
 DT 20-AUG-2001 (Rel. 40, Created)
 DT 20-AUG-2001 (Rel. 40, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE PROBABLE PROTEASE HTPX HOMOLOG (EC 3.4.24.-).
 GN HTPX OR JHP0861.
 OS Helicobacter pylori J99 (Campylobacter pylori J99).
 OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
 OC Helicobacter.
 OX NCBI_TaxID=85963;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA MEDLINE=99120557; PubMed=9923682;
 RA Alm R.A., Ling L.-S.L., Moir D.T., King B.L., Brown E.D., Doig P.C.,
 RA Smith D.R., Noonan B., Guild B.C., deJonge B.L., Carmel G.,
 RA Tummino P.J., Caruso A., Uria-Nickelsen M., Mills D.M., Ives C.,
 RA Gibson R., Merberg D., Mills S.D., Jiang Q., Taylor D.E., Voyis G.F.,
 RA Trust T.J.;
 RT "Genomic sequence comparison of two unrelated isolates of the human
 RT gastric pathogen Helicobacter pylori.";
 RL Nature 397:176-180(1999).
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M48 (ZINC
 CC METALLOPROTEASE).
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial

CC entries requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AE001516; AAD06444.1;
 DR InterPro: IPR001915; Peptidase_M48.
 DR InterPro: IPR000130; zn_mtpeplase.
 DR Pfam: PF01435; Peptidase_M48; 1.
 DR PROSITE: PS00142; ZINC_PROTEASE; FALSE_NEG.
 KW Transmembrane: Hydrolase; Metalloprotease; zinc; Complete proteome.
 FT TRANSMEM 16 36 POTENTIAL.
 FT TRANSMEM 55 75 POTENTIAL.
 FT TRANSMEM 182 202 POTENTIAL.
 FT TRANSMEM 214 234 POTENTIAL.
 FT METAL 166 166 ZINC (CATALYTIC) (BY SIMILARITY).
 FT ACT_SITE 167 167 BY SIMILARITY.
 FT METAL 170 170 ZINC (CATALYTIC) (BY SIMILARITY).
 SQ SEQUENCE 310 AA; 35470 MW; 4188493E1D3C6395 CRC64;

Query Match 6.4%; Score 150.5; DB 1; Length 310;
 Best Local Similarity 23.38; Pred. No. 0.00058;
 Matches 67; Conservative 57; Mismatches 106; Indels 57; Gaps 11;

DB 184 LKIDKPTDFLWYIMVFLVVO--ILAMTIIPV--IMPENKFTPLEDGLKKSISLA 240
 DB 50 LMTQIIFPT-----ITYMFVAVAYIILVCIQNSSIMLSDEKLDPSKVLSSKENQI 104
 OY 241 DRVGFPL-----DKIFVIDGSKSSHSNAVETGIPFSKRIVLFDTLVNSNSTDEI 291
 DB 105 HRLLELLEAKLHMEPRKLYIIN---APYNNAFASGDESNILALTSALIERLDREL 160
 OY 292 TAVLAHEIGHQKNHIVNYIFSQHTFLIFSLFSTIRNTPNTKEFEL--EKSGSF 349
 DB 161 KAVIAHELSHIRHDIRLTMCGILSNIMLVANFSVY-----FFMGNRKNSGAN 210
 OY 350 VDPVITKEPFIILGFMLENDLPLECAMQFVMSLISFHEYQADAVAKKIGY-KONICR 408
 DB 211 LARMLWVLOIILPFLT-----LLQWTLSTREYMDSCAFLMHNKPAIR 258
 OY 409 ALIDL-----QIKNLSTMNVDPLYSSYHS-HPTLAERSTAL 444
 DB 259 ALQKISNDYTNNDYKEIDKNSTSAAYLAFNMESTHRSIKNRISL 305

RESULT 7
 HTPX_ARCFU STANDARD; PRT; 335 AA.
 AC 030004:
 DT 15-JUL-1999 (Rel. 38, Created)
 DT 15-JUL-1999 (Rel. 38, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE PROBABLE PROTEASE HTPX HOMOLOG (EC 3.4.24.-).
 GN HTPX OR AF0235.
 OS Archaeoglobus fulgidus.
 OC Archaeae; Euryarchaeota; Archaeoglobales; Archaeoglobaceae;
 OC Archaeoglobus.
 OX NCBI_TaxID=2234;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA STRAIN=VC-16 / DSM 4304 / ATCC 49558;
 RX MEDLINE=98049343; PubMed=9389475;
 RA Kleink H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,
 RA Ketchum K.A., Dodson R.J., Gwinn M., Hickey E.K., Peterson J.D.,
 RA Richardson D.L., Kerlavage A.R., Graham D.E., Kyriakides N.C.,
 RA Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,
 RA Kirkness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,
 RA Peterson S., Reich C.T., McNeil L.K., Badger J.H., Glodek A., Zhou L.,
 RA Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Utterback T.,
 RA Cotton M.D., Spriggs T., Artlich P., Keine B.P., Sykes S.M.,
 RA Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,
 RA Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,
 RA Venter J.C.;
 RT "The complete genome sequence of the hyperthermophilic, sulphate-

RT reducing archaeon *Archaeoglobus fulgidus*.
 RL Nature 390:364-370(1997).
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M48 (ZINC METALLOPROTEASE).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: AE001090; AAB90998.1; -
 DR MEROPS: M48.013; -
 DR TIGR: AF0235; -
 DR InterPro: IPR001915; Peptidase_M48.
 DR InterPro: IPR000130; Zn_MTPeptide.
 DR Pfam: PF01435; Peptidase_M48; 1.
 DR PROSITE: PS00142; ZINC_PROTEASE; FALSE NEG.
 KW Transmembrane; Hydrolase; Metalloprotease; zinc; Complete proteome.
 FT TRANSMEM 9 29 POTENTIAL.
 FT TRANSMEM 43 63 POTENTIAL.
 FT TRANSMEM 64 84 POTENTIAL.
 FT TRANSMEM 179 199 POTENTIAL.
 FT TRANSMEM 213 233 POTENTIAL.
 FT METAL 168 168 ZINC (CATALYTIC) (BY SIMILARITY).
 FT ACT_SITE 169 169 BY SIMILARITY.
 FT METAL 172 172 ZINC (CATALYTIC) (BY SIMILARITY).
 FT SEQUENCE 335 AA; 37855 MW; 94103E840A2AC68B CRC64;
 SQ
 Query Match 6.4%; Score 150; DB 1; Length 335;
 Best Local Similarity 21.8%; Pred. NO. 0.00069;
 Matches 71; Conservative 51; Mismatches 113; Indels 90; Gaps 13;
 QY 163 TDMIKSLTAVAIIGPILYFLK--IFDKPFDLWYIMVLFVQIILAMTIIIVFIMP 219
 DB 44 TSMVLACMILAIISAIIIXLILAVGVISF-----YGLIFLLINLIMLYLSTYIINL 98
 QY 220 MFNKTPLLEDGLKKSISLADRVGFPDLKIFVIDGSKRSSHNNYFTGLPETSRIYLF 279
 DB 99 SYG--AORDERLQWVSVARLVNKKPKAVV-----RSPNNAFVAGNFTGKFAVVS 150
 QY 280 DTVVNSNSTDEITAVLAHEIGHMKNHTVNNVIFSQLHTPLIFESFTSYNRTSYNFG 339
 DB 151 ESLMMLSQBELAVIGHIGH--HKHNDNAV-----LLEGLLPSTV--FY--LG 195
 QY 340 FLEKST-----GSFVDPVITKEPPIIGMLFNDLTPLECAMQFVMSLISRTHEYOA 393
 DB 196 YALLHSSMRDRRGAQALAIQIA--AVIVSFI-----VQILVAFSRLREYVA 241
 QY 394 DAYAKKLYKONLCALIDL-----QI 415
 DB 242 DFEGRATNKAMORSIAKIHLEFYHRYPDYLAIDSKRFTLFYVATNAVAEPTIRADI 301
 QY 416 KNLSTMANVDPVLYSSYHSHPTLAER 440
 DB 302 EALKMKVSPV--QEFLLSTHPLPKR 325
 RESULT 8
 HTPX_HAEIN STANDARD: PRT: 283 AA.
 AC P44840; HTPX_HAEIN
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE PROBABLE PROTEASE HTPX (EC 3.4.24.-).
 GN HTPX OR HT0720.
 OS Haemophilus influenzae.
 OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;

CC Haemophilus.
 OX NCBI_TaxID=727;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=RD / KW20 / ATCC 51907;
 RX MEDLINE=95350630; PubMed=7542800;
 RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
 RA Kierulff A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
 RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
 RA Scott J.D., Shiley R., Liu L.-I., Glodek A., Kelley J.M.,
 RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
 RA Ueberlack T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
 RA Fife L.D., Fritchman J.L., Fuhrmann J.L., Geoghegan N.S.M.,
 RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
 RA Venter J.C.;
 RT "Whole-genome random sequencing and assembly of Haemophilus
 RT influenzae Rd.";
 RL Science 269:496-512(1995).
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER MEMBRANE
 CC (PROBABLE).
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M48 (ZINC
 CC METALLOPROTEASE).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: U32755; AAC22378.1; -
 DR MEROPS: M48.014; -
 DR TIGR: H10720; -
 DR InterPro: IPR001915; Peptidase_M48.
 DR InterPro: IPR000130; Zn_MTPeptide.
 DR Pfam: PF01435; Peptidase_M48; 1.
 DR PROSITE: PS00142; ZINC_PROTEASE; 1.
 KW Transmembrane; Inner membrane; Hydrolase; Metalloprotease; zinc;
 KW Complete proteome.
 FT TRANSMEM 4 24 POTENTIAL.
 FT TRANSMEM 33 53 POTENTIAL.
 FT TRANSMEM 147 167 POTENTIAL.
 FT TRANSMEM 190 210 POTENTIAL.
 FT METAL 139 139 ZINC (CATALYTIC) (BY SIMILARITY).
 FT ACT_SITE 140 140 BY SIMILARITY.
 FT METAL 143 143 ZINC (CATALYTIC) (BY SIMILARITY).
 FT SEQUENCE 283 AA; 30780 MW; FA0652B1BE0E0818 CRC64;
 SQ
 Query Match 6.3%; Score 147.5; DB 1; Length 283;
 Best Local Similarity 26.5%; Pred. NO. 0.00086;
 Matches 73; Conservative 42; Mismatches 101; Indels 59; Gaps 12;
 QY 198 IMVLEFVQIILAMTIPVIMPENKFTPLD--DGE-----LKKSISSLADRV 243
 DB 37 IMALLFG--FAGSLISFL-----SKTVALNSVGEVITQPRQNTERLIDTVSRQAKA 89
 QY 244 GFDPLKIVIDGSKRSSHNNYFTGLPETSRIYLFDTLVNSNSTDEITAVLAHEIGHM 303
 DB 90 GIPMDVAIV--HSPDVNAFATGATKNSLVAVSTGLNNMTAEAEVLAHEISHIS 145
 QY 304 KNIIVNVIYF--SOLHTPLIF-----SLTSTSYRNTSFYNGFLEKSTGSFVDPVITKE 357
 DB 146 NGDMWYKALLOGVLTAVTFVIFLSRYIATRVASSRNNGEET-----RSSGIYLVSVYLE 199
 QY 358 FPIIGFMLFNDLTPLECAMQFVMSLISRTHEYQADAYAKKLYKONLCALIDLQ--- 414
 DB 200 -----MLFVLAISII--AMWF-----SRREFRADAGSASLVGKEMMALQRLQDLH 245
 QY 415 -----KNLSTMANVDPVLYSSYHSHPTLAERSTAL 444
 DB 246 EPQNLGSLNFMINGKRSLEFMSHPLEKRIEAL 280

```

RESULT 9
HTPX_HELPY
ID HTPX_HELPY STANDARD: PRT: 310 AA.
AC 025582:
DT 15-JUL-1999 (Rel. 38, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE PROBABLE PROTEASE HTPX HOMOLOG (EC 3.4.24.-).
GN HTPX OR HP0927.
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC Helicobacter.
OX NCBI_TaxID=210;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=26695 / ATCC 700392;
RX MEDLINE=97394467; PubMed=9252185;
RA Tomb J.-F., White O., Kervatage A.R., Clayton R.A., Sutton G.G.,
RA Fleischmann R.D., Ketchum K.A., Klenk H.-P., Gill S., Dougherty B.A.,
RA Nelson K., Quackenbush J., Zhou L., Kirkness E.F., Peterson S.,
RA Loftus B., Richardson D., Dodson R., Khalak H.G., Glodek A.,
RA McEnney K., Fitzgerald L.M., Lee N., Adams M.D., Hickey E.K.,
RA Berg D.E., Gocayne J.D., Utterback T.R., Peterson J.D., Kelley J.M.,
RA Cotton M.D., Weidman J.M., Fujii C., Bowman C., Wathey L., Wallin E.,
RA Hayes W.S., Borodovsky M., Karp P.D., Smith H.O., Fraser C.M.,
RA Venter J.C.;
RT "The complete genome sequence of the gastric pathogen Helicobacter
RT pylori."
RL Nature 388:539-547(1997).
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M48 (ZINC
CC METALLOPROTEASE).
CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch.)
CC -----
DR EMBL: AE000602; AAD07972.1; ALT_INIT.
DR MEROPS: M48.014; -.
DR TIGR: HP0927; -.
DR InterPro: IPR001915; Peptidase_M48.
DR InterPro: IPR000130; Zn_MTPeptide.
DR Pfam: PF01435; Peptidase_M48; 2.
DR PROSITE: PS00142; ZINC_PROTEASE; FALSE_NEG.
KW Transmembrane; Hydrolase; Metalloprotease; Zinc; Complete proteome.
FT TRANSMEM 16 36 POTENTIAL.
FT TRANSMEM 55 75 POTENTIAL.
FT TRANSMEM 182 202 POTENTIAL.
FT TRANSMEM 214 234 POTENTIAL.
FT METAL 166 166 ZINC (CATALYTIC) (BY SIMILARITY).
FT ACT_SITE 167 167 BY SIMILARITY.
FT METAL 170 170 ZINC (CATALYTIC) (BY SIMILARITY).
SQ SEQUENCE 310 AA; 35431 MW; 71F909AFC748F3EF CRC64;

Query Match 6.1%; Score 142.5; DB 1; Length 310;
Best Local Similarity 22.7%; Pred. No. 0.0022;
Matches 65; Conservative 55; Mismatches 11; Indels 55; Gaps 10;

OY 184 LKIFKPFPTDFLWYIMVFLVQVLAAMTIIPEF--IMPENKFTPELEGELKSTESLAD 241
DB 50 LMTFOIFPT---IMMIFVAFVIVVVICIONFSSIMSGCKYKILDIKSLKSENGIH 105
OY 242 RVGFPL-----DKIFVIDGSKRSHSNAYFTGLPFTSKRIVLFTLVNSNSTDEIT 292
DB 106 RLLELLEENLHPEPKLYIIN---ADYMNFAFGWNSNLSLIALTALIERLDRDELK 161

```

```

OY 293 AVLAHEIGHMOKNHNIVNVVIESOLHTFLFLSTISYFNTSFYNTFGFEL--EKSTGSFV 350
DB 162 AVIAHELHSHRNDRIRLTMWCGILSNMLLVANSVY-----FFMGRKNSGANL 211
OY 351 DPVITKEPPIITIGFMLFNDLLPLECAMQFVMSLISRTHEYOADAYAKKICV-KONLCRA 409
DB 212 ARMILWQLIILPFLTL-----LLQWLSLRTREYMAVSGAFLHDKKPMIRA 259
OY 410 LIDL-----QIKNLSTMNVDPLSYTHYS-HPLAERSTAL 444
DB 260 LQKISNDYTNNDYKEIDKNSRSAAYLFENAMFSTHPSIKNRISL 305

RESULT 10
HTPX_AERPE
ID HTPX_AERPE STANDARD: PRT: 366 AA.
AC Q9YD67;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE PROBABLE PROTEASE HTPX HOMOLOG (EC 3.4.24.-).
GN HTPX OR APE1045.
OS Aeropyrum pernix.
OC Archaea; Crenarchaeota; Desulfurococcales; Desulfurococcaceae;
OC Aeropyrum.
OX NCBI_TaxID=56636;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K1;
RX MEDLINE=99310339; PubMed=10382966;
RA Kawarabayashi Y., Hino Y., Horikawa H., Yamazaki S., Hatakeyama Y.,
RA Jin-no K., Takahashi M., Sekine M., Baba S.-I., Anka A., Kosugi H.,
RA Hosoyama A., Fukui S., Nagai Y., Nishijima K., Nakazawa H.,
RA Takamizawa M., Masuda S., Funahashi T., Tanaka T., Kudoh Y.,
RA Yamazaki J., Kushiida N., Oguchi A., Aoki K.-I., Kubota K.,
RA Nakamura Y., Nomura N., Sako Y., Kikuchi H.;
RT "Complete genome sequence of an aerobic hyper-thermophilic
RT crenarchaeon, Aeropyrum pernix K1."
RL DNA Res. 6:83-101(1999).
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M48 (ZINC
CC METALLOPROTEASE).
CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch.)
CC -----
DR EMBL: AP000060; BAA80030.1; -.
DR MEROPS: M48.004; -.
DR InterPro: IPR001915; Peptidase_M48.
DR InterPro: IPR000130; Zn_MTPeptide.
DR Pfam: PF01435; Peptidase_M48; 1.
DR PROSITE: PS00142; ZINC_PROTEASE; FALSE_NEG.
KW Transmembrane; Hydrolase; Metalloprotease; Zinc; Complete proteome.
FT TRANSMEM 4 24 POTENTIAL.
FT TRANSMEM 50 70 POTENTIAL.
FT TRANSMEM 79 99 POTENTIAL.
FT TRANSMEM 193 213 POTENTIAL.
FT TRANSMEM 229 249 POTENTIAL.
FT METAL 181 181 ZINC (CATALYTIC) (BY SIMILARITY).
FT ACT_SITE 182 182 BY SIMILARITY.
FT METAL 185 185 ZINC (CATALYTIC) (BY SIMILARITY).
SQ SEQUENCE 366 AA; 39742 MW; 3DF2AF596CC67584 CRC64;

Query Match 5.9%; Score 139; DB 1; Length 366;
Best Local Similarity 22.2%; Pred. No. 0.0047;
Matches 74; Conservative 52; Mismatches 13; Indels 74; Gaps 12;

```

```

OY 169 LTIAAIGGPIYLLELKF--DKPTDGLMYLMVLFVQIQAAMTII-----PFIPIWE 221
Db 50 LTVALLIGGIYIVLVGAAMALDAF-TGLSGTCVAPMLSLALAMVLQMLSPYIIMNVY 108
OY 222 NKFTPLEDGE-LKKSIESLADRVGFPDLKIEYIDGSKRSSHNAVFTGLPFTSKRIYLD 280
Db 109 RTREPLPEEEMVIAVEEVLQARRSGLKPKRVVV-----SEMMNPNAPVSGPIGSAVATR 164
OY 281 TLVYNSNIDETAVLAHLEIGHQKHIIYNNMVFISQHLHFLIFLSFTSYIRNTSPFNTIGF 340
Db 165 GLRLLPKDEYNAVLAHEVGH-LKHRDVTVLIALSLPIAAF---LIGRLVWAGILGG 219
OY 341 FLEKSTGSEVDPVITKEPPIIGFIMFLFNDLLTPLECAMQFVSLISRTHEQADAY-AKK 399
Db 220 GGGERGRGNPM-ALVAVGALLAGVNF-----QLYSHENRLREYIADHSALV 267
OY 400 LGYKONLCRAL-----IDLOIKN 417
Db 268 TGSPRSRLQALARIHAAYEHNPMLVEEARSNEMASMLFVAPLTSLTASPLVDVQYLYER 327
OY 418 LSTNAVDPPIYSYHSHPPLAERSALDQYSK 450
Db 328 LKEQETNPVLVELS-TTPPVSKRLRFELDLRASR 359

RESULT 11
HTPX_METTH
ID HTPX_METTH STANDARD: PRT: 258 AA.
AC 026659;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE PROBABLE PROTEASE HTPX HOMOLOG (EC 3.4.24.-).
GN HTPX OR MTH569.
OS Methanobacterium thermoautotrophicum.
OC Archaeae; Euryarchaeota; Methanobacteriales; Methanobacteriaceae;
OX Methanothermobacter.
NX NCBI_TaxID=145262;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DELTA H;
RX MEDLINE=98037514; PubMed=9371463;
RA Smith D.R., Doucette-Stamm L.A., Deloughery C., Lee H.-M., Dubois J.,
RA Aldredge T., Bashirzadeh R., Blakely D., Cook R., Gilbert K.,
RA Harrison D., Hoang L., Keagle P., Lumu R., Potlter B., Qiu D.,
RA Spadafora R., Viocare R., Wang Y., Wierzbowski J., Gibson R.,
RA Jivani N., Caruso A., Bush D., Saffer H., Patwell D., Prabhakar S.,
RA McConaughy S., Shimer G., Goyal A., Pietrovski S., Church G.M.,
RA Daniels C.J., Mo J.-I., Rice P., Nolling J., Reeve J.N.;
RT "Complete genome sequence of Methanobacterium thermoautotrophicum
RT deltaH: functional analysis and comparative genomics.";
RL J. Bacteriol. 179:7135-7155(1997).
CC -1 SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC -1 SIMILARITY: BELONGS TO PEPTIDASE FAMILY M48 (ZINC
CC METALLOPROTEASE).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/or\_send\_an\_email\_to\_license@isb-sib.ch).
CC -----
CC EMBL: AE000839; AAB85075.1; -
CC MEROPS: M48.004; -
CC InterPro: IPR001915; Peptidase_M48.
CC InterPro: IPR001030; Zn_MTPpeptidse.
CC Pfam: PF01435; Peptidase_M48; 1.
CC PROSITE: PS00142; ZINC_PROTEASE; FALSE_NEG.
KW Transmembrane; Hydrolase; Metalloprotease; Zinc; Complete proteome.
FT TRANSMEM 24 44
FT TRANSMEM 45 65
FT POTENTIAL.

```

| FT | TRANSMEM | 157 | 177 | POTENTIAL. |
|-----------------------|--|-----------------------------------|--|-----------------------------------|
| FT | TRANSMEM | 186 | 206 | POTENTIAL. |
| FT | METAL | 146 | 146 | ZINC (CATALYTIC) (BY SIMILARITY). |
| FT | ACT_SITE | 147 | 147 | BY SIMILARITY. |
| FT | METAL | 150 | 150 | ZINC (CATALYTIC) (BY SIMILARITY). |
| FT | SEQUENCE | 258 AA; | 28605 MM; | DBD9884292ABPAJ3 CRC64; |
| SO | | | | |
| Query Match | | 5.9% | Score 138.5; | DB 1; Length 258; |
| Best Local Similarity | | 22.0%; | Pred. No. 0.0034; | |
| Matches 63; | Conservative | 46; | Mismatches 112; | Indels 65; Gaps 10; |
| QY | 147 | LEERGFN---- | KLTVOIMTDMKSLTLANV----- | GGPLLYLFKTFDKEP 191 |
| DB | 1 | MEKKMKRRRLSTWKLKRLKELATV-L | GLYAIALLMVGSILGGLPYFALL---- | GGF 55 |
| QY | 192 | TDFEYIYVFLFVVOILAMTITP | YIMPMFNKFPLEDEGLKKSIESLADRVGFPLDKIF | 251 |
| DB | 56 | VIFLQY----- | LISPRIVELTMNVHV----- | SEAEAPRLHAMVDELARAGIKPKPV- 103 |
| QY | 252 | VIDGSKRSSHSNAYFTTG | LFTFSKRIYLFDTLVNSSTDEITAVLAHEIGHWOKNIYVNV | 311 |
| DB | 104 | ---GIAEIALPAAFAFG | GRTKADGRVCVIRGILINLDEEELRAVLGHEISHIKHSDMIYMT | 160 |
| QY | 312 | IFSQHTFLFLFLFSIR | FNSTSFYNTPGFPLEKSGSVDPVPIYTEFPILIGFMFLNDLL | 371 |
| DB | 161 | LVSAPVPLICYIYEWSTFVS | SRDDEALVVG----- | IAALLVYFVG----- 198 |
| QY | 372 | TPLECAMQFVMSLISRT | HEYQADAVAKKL-GYKONICRALIDLOJIK | 416 |
| DB | 199 | ----- | OLIVFLFIRTRREYVADQSV | ELGGPHKLAALTKVLVR 237 |
| RESULT 12 | | | | |
| ID | HTPX_PYRHO | STANDARD: | PRT: | 292 AA. |
| AC | 0586957; | | | |
| DT | 15-JUL-1999 | (Rel. 38, Created) | | |
| DT | 15-JUL-1999 | (Rel. 38, Last sequence update) | | |
| DE | 20-AUG-2001 | (Rel. 40, Last annotation update) | | |
| GN | HTPX OR PH1256. | | | |
| OS | Pyrococcus horikoshii. | | | |
| OC | Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus. | | | |
| OX | NCBI_TaxID=53953; | | | |
| RN | [1] | | | |
| RP | SEQUENCE FROM N.A. | | | |
| RC | STRAIN=OT3; | | | |
| RX | MEDLINE=98344137; | PubMed=9679194; | | |
| RA | Kawarabayashi Y., Sawada M., Horikawa H., Haikawa Y., Hino Y., Yamamoto S., Sekine M., Baba S.-I., Kosugi H., Hosoyama A., Nagai Y., Sakai M., Ogura K., Otsuka R., Nakazawa H., Takamiya M., Ohfuku Y., Funahashi T., Tanaka T., Kudoh Y., Yamazaki J., Kushiida N., Oguchi A., Aoki K.-I., Yoshizawa T., Nakamura Y., Robb F.T., Horikoshi K., Masuchi Y., Shizuya H., Kikuchi H.; | | | |
| RA | "Complete sequence and gene organization of the genome of a hyper-thermophilic archaeobacterium, Pyrococcus horikoshii OT3." | | | |
| RT | DNA Res. 5:55-76(1998). | | | |
| RL | -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL). | | | |
| CC | -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M48 (ZINC METALLOPROTEASE). | | | |
| CC | ----- | | | |
| CC | THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch). | | | |
| CC | EMBL, AP000005; BAA30357.1; - | | | |
| DR | MEROPS; M48_004; - | | | |
| DR | InterPro; IPR001915; Peptidase_M48 | | | |

[illegible]

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: X04465; CAA28078.1; -
DR PIR: A05037; A05037.
DR PIR: S01591; S01591.
DR Mendel: 5295; MAPPO.ycf2.1.
DR InterPro: IPR001939; AAA_subfam.
DR Pfam: PF00004; AAA.1.
KW Chloroplast; Hypothetical protein.
SQ SEQUENCE 2136 AA; 259911 MW; 5BD170C9CCF61197 CRC64;

Query Match 5.4%; Score 127.5; DB 1; Length 2136;
Best Local Similarity 21.3%; Pred. No. 0.26;
Matches 113; Conservative 73; Mismatches 171; Indels 173; Gaps 28;

OY 1 MFDLTIIDHPNIP-----WKLIIT-----SGFSIAQSFSESYLTYYKQOKLSETKLPVLE 51
DB 156 LIDRSILKKKEIILHMKFLVLEQIOSNMKISE-----ESLSELTKE--VLE 201
OY 52 DEIDDEPFH-----KSNRYSAKKKFSFGVY 79
DB 202 OKNIDELKHFEFFINOKIYNNMWEYFYISFINQKIDIKSNKYNNSGFEFE--L 258
OY 80 NLAQKLVFIKYDEPKIMHMAVSLNANVLPVRFHNVSTVAOSLCFLGLSLSLVLDPL 139
DB 259 AFCEKLF-----EVEFLSKPNNNNLQMKLNCLENFSLDLF-----CLNNKL 302
OY 140 SYTHFVLEEFKGNKLTQVLMITDMKSLTFLAVALGPGILYLKLFIDKFPDPLWYIM 199
DB 303 PVMKKIKFKNLQNNESDKL-----IESFPLKIKGN-----LYFKNIIEFVT--W--- 347
OY 200 VLEFVOIAMLTIIPVIMPENKFTPLEDEGLKSTESLADRGVFDKIFVIDGS--- 256
DB 348 -----QSYKKDCID-----FNKFNELNSEIYIKIEELSDVIYKFSKILVEGKSK 395
OY 257 ---KRSSHNAVFTGLP--FTSKRIVLEDTLVNSNSTDEITAVLAHEIGHW--QKNHIVNM 310
DB 396 TIINQSFNNIYKKLNSENFENFIYED-----SNNL-----LFDLKKNIYINN 441
OY 311 VIFSQLHFLFLSLTSTYRNTSFYNTF--GFFLEKSTGCF-----VDPVITKE 357
DB 442 KPF--LKSFLIYS-----SISNOFILFFKQNSKSKFNKLVKKNKSDVITNVFSKE 490
OY 358 PFIIG-----FMLENDLPLPLCAMQFVMSLSRTH---EYQADAVAKKLGKONLCR 408
DB 491 NKIEINFSKSIYAFELISINELDKNFYINKISLKNKKOKRFLYLNKIKSSDNF-- 548
OY 409 ALIDL-QIKNLSTMNVDPLVSSYHSHPTLAERSTALDYVSE---KKKN 453
DB 549 RFIULMKIKN-----YSSQFVSNNSFLNPAPEILQONYYLKKKN 589

RESULT 15
HTPX_BUCAL STANDARD; PRT; 292 AA.
AC P57406;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE PROBABLE PROTEASE HTPX HOMOLOG (EC 3.4.24.-).
GN HTPX OR BU321.
OS Buchnera aphidicola (subsp. Acyrthosiphon pisum) (Acyrthosiphon pisum
OC symbiotic bacterium).
OC Bacteria; Proteobacteria; gamma subdivision; Buchnera.
OX NCBI_TaxID=118099;
RN [1]

RP SEQUENCE FROM N.A.
RC STRAIN=OKYO 1998;
RX MEDLINE=20445173; PubMed=10993077;
RA Shigenobu S., Matanabe H., Hattori M., Sasaki Y., Ishikawa H.;
RT "Genome sequence of the endocellular bacterial symbiont of aphids
RL Buchnera sp. Aps. ";
RL Nature 407:81-86(2000).
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M48 (ZINC
CC METALLOPROTEASE).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: AP001119; BAB13029.1; -
DR InterPro: IPR001915; Peptidase_M48.
DR InterPro: IPR00130; Zn_Mtpepdse.
DR Pfam: PF01435; Peptidase_M48.1.
DR PROSITE: PS00142; ZINC_PROTEASE; FALSE_NEG.
KW Transmembrane; Hydrolase; Metalloprotease; Zinc; Complete proteome.
FT TRANSMEM 4 24 POTENTIAL.
FT TRANSMEM 32 52 POTENTIAL.
FT TRANSMEM 147 167 POTENTIAL.
FT TRANSMEM 193 213 POTENTIAL.
FT METAL 139 139 ZINC (CATALYTIC) (BY SIMILARITY).
FT ACT_SITE 140 140 BY SIMILARITY.
FT METAL 143 143 ZINC (CATALYTIC) (BY SIMILARITY).
SQ SEQUENCE 292 AA; 32762 MW; DN0135106794BA30 CRC64;

Query Match 5.1%; Score 120; DB 1; Length 292;
Best Local Similarity 23.1%; Pred. No. 0.083;
Matches 68; Conservative 48; Mismatches 101; Indels 78; Gaps 14;

OY 200 VLEFVOIAMLTIIPVIMP-----MENKFTPLE--DGL-----KSTIES-LA 240
DB 20 LILFLTGIOSTWITVIGLIMSGLFGSGSILSLIKMIALSVNGEITTHRENEVSWLI 79
OY 241 DRV-----GFLDKIFVIDGSKRSSHNAVFTGLPFTSKRIYVFDLVNSNDELTA 293
DB 80 NTVRQOSTOKGIIHQIAVY---HAYDINAFATGARNASALIAVSTGLLENMTHHEEA 135
OY 294 VLAHEIGHMOKNHIYVM--VIFSOLHTELP--SLFTSIY-----RNTSEFYNTGCF 340
DB 136 VIAHEISHIANGDMITMTLVGVVNTFVILISRLSOTIISVMSNRENNTTEKNSFY 195
OY 341 FLEKSTGFDVPVITKEPPIIGFMLENDLPLPLCAMQFVMSLSRTHYEQADAVAKL 400
DB 196 FL---VSTFLE-----LIFGIL-----ASITMWFSSRRREFYADASSAKM 232
OY 401 GYKONLCALDLQIKNLSTMNVDPL-----YSSYHSHPTLAERSTAL 444
DB 233 VGREMTAALRLKTSR--EPOSDSMIAFCINGKSKSFLKLPASHPSLENRIEAL 286

Search completed: February 12, 2002, 09:32:24
Job time: 340 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 12, 2002, 09:25:34 ; Search time 26.02 Seconds

(without alignments)
1326.175 Million cell updates/sec

Title: US-09-165-460a-2

Perfect score: 2342

Sequence: 1 MFDLKTIIDHPNIPMKLIIS.....HPTLAERSTALDYSEKKKN 453

Scoring table: BLOSUM62

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB | ID | Description |
|------------|-------|-------------|--------|----|--------|--------------------|
| 1 | 2336 | 99.7 | 453 | 2 | S57140 | probable membrane |
| 2 | 903 | 38.6 | 474 | 2 | T38737 | probable CAAX pren |
| 3 | 819.5 | 35.0 | 459 | 2 | C85017 | probable CAAX pren |
| 4 | 654 | 27.9 | 316 | 2 | T01712 | hypothetical prote |
| 5 | 612.5 | 26.2 | 442 | 2 | T18917 | hypothetical prote |
| 6 | 501.5 | 21.4 | 395 | 2 | F81343 | probable integral |
| 7 | 459 | 19.6 | 407 | 2 | D71862 | probable zinc-met |
| 8 | 454 | 19.4 | 407 | 2 | F64567 | zinc metalloprotei |
| 9 | 378 | 16.1 | 428 | 2 | E84192 | zinc metalloprotei |
| 10 | 328 | 14.0 | 426 | 2 | C69831 | conserved hypotet |
| 11 | 197 | 8.4 | 338 | 2 | H75127 | CAAX prenyl protei |
| 12 | 152 | 6.5 | 291 | 2 | D83293 | heat shock protein |
| 13 | 150.5 | 6.4 | 310 | 2 | D71878 | stress protein - H |
| 14 | 150 | 6.4 | 335 | 2 | C69279 | heat shock protein |
| 15 | 148.5 | 6.3 | 279 | 2 | E81951 | probable membrane |
| 16 | 148.5 | 6.3 | 347 | 2 | E81154 | heat shock protein |
| 17 | 147.5 | 6.3 | 283 | 1 | H64088 | heat shock protein |
| 18 | 142.5 | 6.1 | 326 | 1 | G64635 | heat shock protein |
| 19 | 139 | 5.9 | 366 | 1 | F72703 | probable heat shoc |
| 20 | 138.5 | 5.9 | 258 | 1 | C69175 | heat shock protein |
| 21 | 136 | 5.8 | 287 | 2 | E82239 | heat shock protein |
| 22 | 135.5 | 5.8 | 289 | 2 | B82334 | heat shock protein |
| 23 | 130.5 | 5.6 | 292 | 2 | C71070 | probable heat shoc |
| 24 | 129.5 | 5.5 | 304 | 2 | F83987 | heat-shock protei |
| 25 | 127.5 | 5.4 | 2136 | 2 | A05037 | hypothetical prote |
| 26 | 123 | 5.3 | 315 | 2 | E69365 | hypothetical prote |
| 27 | 122 | 5.2 | 3119 | 2 | T18414 | protein g377 - mal |
| 28 | 120.5 | 5.1 | 289 | 2 | A75094 | heat shock protein |
| 29 | 120 | 5.1 | 292 | 2 | E84967 | heat shock protein |

| | | | | | | |
|----|-------|-----|------|---|--------|--------------------|
| 30 | 119 | 5.1 | 297 | 2 | T48855 | probable heat shoc |
| 31 | 117.5 | 5.0 | 324 | 2 | B86672 | permease (imported |
| 32 | 116 | 5.0 | 293 | 1 | A43659 | heat shock protein |
| 33 | 116 | 5.0 | 293 | 1 | G83794 | heat shock protein |
| 34 | 113 | 4.8 | 284 | 1 | H64509 | heat shock protein |
| 35 | 113 | 4.8 | 298 | 1 | G69862 | heat-shock protein |
| 36 | 110.5 | 4.7 | 355 | 2 | T29251 | hypothetical prote |
| 37 | 109.5 | 4.7 | 312 | 2 | G70114 | phosphatidate cyti |
| 38 | 109.5 | 4.7 | 669 | 2 | S76474 | hypothetical prote |
| 39 | 109 | 4.7 | 506 | 2 | S34457 | hypothetical prote |
| 40 | 109 | 4.7 | 595 | 2 | E64460 | hypothetical prote |
| 41 | 107.5 | 4.6 | 1032 | 2 | S53571 | hypothetical prote |
| 42 | 106.5 | 4.5 | 645 | 2 | T41876 | p74 orf138 - Bomby |
| 43 | 106.5 | 4.5 | 1029 | 2 | S56229 | probable membrane |
| 44 | 106 | 4.5 | 842 | 2 | T04880 | potassium transpor |
| 45 | 105.5 | 4.5 | 478 | 2 | T11318 | MDH dehydrogenase |

ALIGNMENTS

RESULT 1

S57140 probable membrane protein YJRI17w - yeast (Saccharomyces cerevisiae)

N: Alternate names: hypothetical protein J2032

C: Species: Saccharomyces cerevisiae

C: Date: 23-Aug-1995 #sequence, revision 08-Sep-1995 #text, change 06-Feb-1998

C: Accession: S57140

R: Rose, M.; Koetter, P.; Entian, K.D.

submitted to the Protein Sequence Database, September 1995

A: Accession number: S56848

A: Molecule type: DNA

A: Residues: 1-453 <ROS>

A: Cross-references: EMBL:Z49617; NID:G1015836; PID:G1015837; MIPS:YJRI17w

C: Genetics:

A: Gene: SGD:STE24

A: Cross-references: SGD:S0003878; MIPS:YJRI17w

A: Map position: 10R

C: Keywords: transmembrane protein

| | | | | |
|-----------------------|-----------------|--|-----------|-------------|
| Query Match | 99.7% | Score 2336; | DB 2; | Length 453; |
| Best Local Similarity | 99.8% | Pred. No. 4.9e-171; | | |
| Matches 452; | Conservative 0; | Mismatches 1; | Indels 0; | Gaps 0; |
| QY | 1 | MFDLKTIIDHPNIPMKLIISGFSIAQSFESYLYRYOKLSEKLPVLEDEIDDETFH | 60 | |
| DB | 1 | MFDLKTIIDHPNIPMKLIISGFSIAQSFESYLYRYOKLSEKLPVLEDEIDDETFH | 60 | |
| QY | 61 | KSRNYSRAKAKFSIFGDVYNLAOKLVPFKYDLFPKIMHMAVSLNAVLPVRFHNVSTVAQ | 120 | |
| DB | 61 | KSRNYSRAKAKFSIFGDVYNLAOKLVPFKYDLFPKIMHMAVSLNAVLPVRFHNVSTVAQ | 120 | |
| QY | 121 | SLCFLGLISLSTVVDLPVLSYSHFVLEKSGPNKLTQVLTMTMISLTLAVYAGPIL | 180 | |
| DB | 121 | SLCFLGLISLSTVVDLPVLSYSHFVLEKSGPNKLTQVLTMTMISLTLAVYAGPIL | 180 | |
| QY | 181 | YLFLEKIDKPTDFLWYIMVFLFVQILAMTIPFIMPMKFKTPLEDGELKKSIESLA | 240 | |
| DB | 181 | YLFLEKIDKPTDFLWYIMVFLFVQILAMTIPFIMPMKFKTPLEDGELKKSIESLA | 240 | |
| QY | 241 | DRVGFPLDKIVIDGSKRSHSNAYFGLPPTSRIYVFDPLVNSNSTDETTAVLAHEIG | 300 | |
| DB | 241 | DRVGFPLDKIVIDGSKRSHSNAYFGLPPTSRIYVFDPLVNSNSTDETTAVLAHEIG | 300 | |
| QY | 301 | HMOKNHIYNAVYFSQLHFLFSLFSTSYRNTSPYNTGPFLEKSTGSEFVDPVTKKEPI | 360 | |
| DB | 301 | HMOKNHIYNAVYFSQLHFLFSLFSTSYRNTSPYNTGPFLEKSTGSEFVDPVTKKEPI | 360 | |
| QY | 361 | TIGFMFLNDLTLPLECAMQFVMSLISRTHEYOADAAYAKKLGKONLCRALIDLIQKNLST | 420 | |
| DB | 361 | TIGFMFLNDLTLPLECAMQFVMSLISRTHEYOADAAYAKKLGKONLCRALIDLIQKNLST | 420 | |

OY 421 MNVDLYSSHYSHPTLAERSTALDYSEKKKN 453
 |||||
 Db 421 MNVDLYSSHYSHPTLAERSTALDYSEKKKN 453

RESULT 2

T38737
 probable CAAX prenyl proteinase 1 - fission yeast (Schizosaccharomyces pombe)

C:Species: Schizosaccharomyces pombe
 C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 07-Dec-1999
 R:Accession: T38737
 R:Genetic: S. J. Church, C. M. J. Barrett, B. G. Rajandream, M. A. Walsh, S. V. submitted to the EMBL Data Library, December 1995
 A:Reference number: 221808

A:Accession: T38737
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-474 <GEN>
 A:Cross-references: EMBL:Z68144; PIDN:CAA92258.1; GSPDB:GN00066; SPDB:SPAC3H1.05
 A:Experimental source: strain 972h-; cosmid c3H1
 A:Gene: SPDB:SPAC3H1.05
 A:Map position: 1
 A:Introns: 35/3

Query Match 38.6%; Score 903; DB 2; Length 474;
 Best Local Similarity 40.4%; Pred. No. 176-61;

Matches 184; Conservative 96; Mismatches 153; Indels 22; Gaps 6;

OY 4 LKTHIDHPNIMKLIISGFSIAQSFESYLYROYOKLSEKLPVLEDEIDDEFHRSR 63
 |||||
 Db 37 LMHTIDGFPKVIAGSFGKAYMDLYLRQVRYLLREKPAIILAEHDEKKYQAL 96
 OY 64 NYSRAKAFSIFGDVYNLAQKLVFKYDLPKIMHMA---VSLNAVLPVRFHNVSTVA 119
 |||||
 Db 97 SYARDKSWFSTYVSTFLAVDLILIKYDLSYLMNITKFPMMDKL-AASSRFSLSSTIT 155
 OY 120 QSLCEFLGLSLTLVDLPYSYSHFVLEKFGFNKLTVOIMTDMKSLTAVAGPI 179
 |||||
 Db 156 HSCVFMELTFLSRLIQIPNLSTFVLEKFGFNKLTVOIMTDMKSLTAVAGPI 215
 OY 180 LYLFLKIDKPTDPLAVIMVFLFVVOILAMTIIIPVFMPEKFTPLEDEGLKKSISL 239
 |||||
 Db 216 VGVVFKILTKGDNFIMKAGAYVFGILDTIAPSLIMPLRYKFTPLENSLKTQIDEL 275
 OY 240 ADRVGFPDLKIFVIDGSRSSSHNAYFTGLPFTSKRIYLPDTLVNSSTDEITAVLAHEI 299
 |||||
 Db 276 AASINFPKRLKYVIDASRSTHNAFVGLPW-NKGIYLPDTLVKNHPEPELIALIGHEL 334
 OY 300 GHMOKNH-IVNMVIFQSOLHTFLISLFTSIRNTSFYTFGFELEKSGSFVDPIYTKEF 358
 |||||
 Db 335 GHMWSHNLMTIIDYGSFLHLP-LFAAFIRNNSLSTYSFNTFEK----- 379
 OY 359 PIITGFMFLNDLTPLECAMQFVMSLSRTHGYOADAFAKKGYNKCRALIDLOIKNL 418
 |||||
 Db 380 PVIYGLLFSDALGPLSSILTFASKNYSRLCEYQADAPAKQGLGDLIRIHNDNL 439
 OY 419 STMNVDPYSSHYSHPTLAERSTALDYSEKKKN 453
 |||||
 Db 440 SPLEFDSLSTYSVSHPTLVDRNALIDYTLTKKN 474

RESULT 3

C85017
 probable CAAX prenyl proteinase [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)
 C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 16-Feb-2001

R:Accession: C85017
 R:Anonymous, The European Union Arabidopsis Genome Sequencing Consortium, The Cold Spring Nature 402, 769-777, 1999
 A:Title: Sequence and analysis of chromosome 4 of the plant Arabidopsis thaliana.

A:Reference number: A85001; MUID:20083488

A:Accession: C85017
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-459 <STO>
 A:Cross-references: GB:NC_001268; NID:g7267629; PIDN:CAB80941.1; GSPDB:GN00140
 C:Genetics:
 A:Gene: AT4G01320
 A:Map position: 4

Query Match 35.0%; Score 819.5; DB 2; Length 459;
 Best Local Similarity 38.1%; Pred. No. 46-55;

Matches 184; Conservative 73; Mismatches 155; Indels 71; Gaps 9;

OY 13 IPWKLIISGFSIAQSFESYLYROYOKLSEKLPVLEDEIDDEFHRSRYSRAKAF 72
 |||||
 Db 3 IPFMTVVGFMIVMVFETVLDRLQTLKPLTLEKTLGVISQKFEKSRAYR----- 56
 OY 73 SIFGDVYNLAQKLVFK-----YDLPKIMHMAVSLNAVLP-VRFHNVSTV 118
 |||||
 Db 57 DITENFNICSYFHFVHEPVTIIMDSALFPGILFWMKMS---GAVLPRLGLDPEMET 112
 OY 119 AQSICFLGLSLTLVDLPYSYSHFVLEKFGFNKLTVOIMTDMKSLTAVAGPI 178
 |||||
 Db 113 LHTLSFLAGVMTWSQITDLPESLYSTFVIESRHGFKQTIMFIMDMKGFSLYLGP 172
 OY 179 ILVFLKIDKPTDPLAVIMVFLFVVOILAMTIIIPVFMPEKFTPLEDEGLKKSIS 238
 |||||
 Db 173 IVAALIFVOKGPPYLAIVLWAFMFLSLVMVITVPVLAFLFNKFTPLPGDDREKTEK 232
 OY 229 LADRVGFPDLKIFVIDGSRSSSHNAYFTGLPFTSKRIYLPDTLVNS-NSTDEITAVLAH 297
 |||||
 Db 223 LASSIKFPLKLFVVDGSTRSSSHNAYFG-FNKRRLVLDTLQCKNDEITAVLAH 291
 OY 298 EIGHMOKNHIVNMVIFQSOLHTFLISLFTSIRNTSFYTFGFELEKSGSFVDPIYTK 357
 |||||
 Db 292 ELGHMKNLHTTYSFIAVQILAFLOFGGYTLVRNSDTLRSRSGFDQ----- 337
 OY 338 PPIITGFMFLNDLTPLECAMQFVMSLSRTHGYOADAFAKKGYNKCRALIDLOIK- 416
 |||||
 Db 338 -PVLIGLIIIFQHTVPIPLHVSFGLNVSRAFEFQADAFVAKGLRPLVXLYQRE 396
 OY 417 -----NLSTMNVDPYSSHYSHPTLAERSTALDYSE 449
 |||||
 Db 397 DNNRTQYTSICVTHLNGFVGILOENLSAMNDPLSAHYSHPTLVERLAID--GE 454
 OY 450 KKK 452
 |||||
 Db 455 DKK 457

RESULT 4

T01712
 hypothetical protein A.IG002N01.21 - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)
 C:Date: 19-Feb-1999 #sequence_revision 19-Feb-1999 #text_change 24-Mar-1999

R:Accession: T01712
 R:Schaefer, P.; Magg, L.
 submitted to the EMBL Data Library, June 1997
 A:Description: The sequence of A. thaliana IG002N01.
 A:Reference number: 214407

A:Accession: T01712
 A:Status: translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-316 <SCH>
 A:Cross-references: EMBL:AF007269; NID:g2191126; PID:g2191141
 C:Genetics:
 A:Experimental source: cultivar Columbia
 A:Map position: 4
 A:Introns: 11/1; 40/3; 62/3; 95/3; 133/3; 170/3; 222/3; 246/3; 278/3
 A:Note: A_IG002N01.21

Query Match 27.9%; Score 654; DB 2; Length 316;
 Best Local Similarity 34.1%; Pred. No. 1.1e-42;
 Matches 154; Conservative 54; Mismatches 93; Indels 150; Gaps 9;

QY 13 IPWKIIIGFSIAQSFESYLTROYOKLSETKLPVLEDEIDETFKSRNYSRAKAF 72
 DB 3 IPFMETVVGFMIVMYIFETYLDRQ---LFALEKLPPL----- 36

QY 73 SIFGVYVLAQKLVIRIKIDLFPKIMHMAVSLNANLVPFRHNVSTVAOSLCFLGLLSLS 132
 DB 37 -----PK-----T 39

QY 133 TLVDPLSYSHFVLEKFGFNKLTVOIMTDMIKSLTAVAGGPIILYFLKIPDKPPT 192
 DB 40 LITDLPFLSYSTFVIESHNGKOTIMFIRMIGTFLSYLGGPIYAAITFIYQKGP 99

QY 193 DFLWYVWFLVVOILMTIIPVIMPENKFTPLEDELKKSIESLADRVGFPLDKITFV 252
 DB 100 YLAITLMAFMFILTSLVMTIYVPLIAPLKNKFTPLPDGDLREKIEKLASLKFLPKLFV 159

QY 253 IDGSRSSHNAAYFTGLPFTSKRIYLFETLVNS-NSTDEITAVLAHEIGHMOKNIIVMV 311
 DB 160 VDGSTRSSHNAAYMGF-FKKRKRIYLYDPTLIQOCKNEDEIYAVLAHEIGHMOKNIH- 213

QY 312 IFSOLHTFLIFSLFTSIYRNTSFYNTFGFLEKSTGSPVDPYITKEPFIIGFMFLNDL 371
 DB 214 -----TTSFIAVQHT-----V 225

QY 372 TPLECAMQVMSLISRTHEYQADAVAKKLYKONICRALIDLOK-----NLSTM 421
 DB 226 IPLHLVSGFLNLSVRAEFQADAFVAKKLYKONICRALIDLOK-----NLSTM 285

QY 422 NVDPVLSYSHYSHPTLAERSTALDYSEKK 452
 DB 286 NTDPLYSAYHSHPELVERLAID--GDDK 314

RESULT 5
 T18917
 hypothetical protein C04F12.10 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
 C:Accession: T18917
 R:Lloyd, C.
 Submitted to the EMBL Data Library, November 1996
 A:Reference number: Z19044
 A:Accession: T18917
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-442 <WIL>
 A:Cross-references: EMBL:Z81461; PIDN:CA803839.1; GSPDB:GN00019; CESP:C04F12.10
 A:Experimental source: clone C04F12
 A:Gene: CESP:C04F12.10
 A:Map position: 1
 A:Introns: 30/2; 77/3; 318/1; 340/3

Query Match 26.2%; Score 612.5; DB 2; Length 442;
 Best Local Similarity 32.1%; Pred. No. 2.5e-39;
 Matches 150; Conservative 82; Mismatches 155; Indels 81; Gaps 10;

QY 25 AQFSESLTYROYO-KLSETKLPVLEDEIDETFKSRNYSRAKAFSIFGDVYNLAQ 83
 DB 16 ALFLMDQITFRQYKAHNAVAKRPEVVELIGDEEDYKARQKIDNHLGFFHSWFNQL 75

QY 84 KLVEIKYDLFPKIMHMAVSLNANLVPFRHNVSTVAOSLCFLGLLSSTLVDPVLSYS 143
 DB 76 LTAQIGYGFYFWMTATSS-----YPLHVA-----VLSINISITETIIDLPWDLXS 121

QY 144 HFLVEEKFGFKLTVOIMTDMIKSLTAVAGGPIILYFLKIPDKPPTDPLWYIMVLF 203
 DB 144 HFLVEEKFGFKLTVOIMTDMIKSLTAVAGGPIILYFLKIPDKPPTDPLWYIMVLF 203

DB 122 TFIIDAHGFKOTIGFVVDKIKKMLVGFALTMPIVYGIEWIVNGSPFYVIMLFVS 181
 QY 204 VVOIILMTIIPVIMPENKFTPLEDELKKSIESLADRVGPPDKIPYIDSKSSSHN 263
 DB 182 VVLLIMTITPFIAPLEFKYPLPDGDLKRTIEQLAASLYPLTELYVNGSKRSASHN 241

QY 264 AFTGLPFTSKRIYLFETLVN-----NSTDEITAVLA 296
 DB 242 AYMGF-WKNRKRIYLYDPTLIQOCKNEDEIYAVLAHEIGHMOKNIIVMV 300

QY 297 HEIGHMOKNI-IVNVIESQILHTFLISLTISYRNTSFYNTFGFLEKSTGSPVDPYIT 355
 DB 301 HELGHMALWHTLINLVI-TEVNLFPSPFAVGYGYWEALYQGFY----- 344

QY 356 KEPIIIGFMFLNDLTPLECAMQVMSL-----SRTHEYQADAVAKKLYK 404
 DB 345 HDTPPVIGMMLT-----FOEVLALYNOLASIGVYHRSASAFGADERANLGHGE 394

QY 405 NLGALIDLOKNIKSLTMVNDPLYSYSHYSHPTLAERSTALDYSEKK 452
 DB 395 NLGALTLGVNLSMPINDSLXMCYTHPPVERVAARAFQAKNK 442

RESULT 6
 F81343
 probable integral membrane zinc-metalloproteinase Cj0723c [imported] - Campylobacter
 C:Species: Campylobacter jejuni
 C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 31-Mar-2000
 C:Accession: F81343
 R:Parkhill, J.; Wren, B.W.; Mungall, K.; Kelley, J.M.; Churcher, C.; Basham, D.; Chli
 C.W.; Quail, M.; Rajadurai, M.A.; Rutherford, K.M.; VanVleet, A.; Whitehead, S.; B
 Nature 403, 665-668, 2000
 A:Title: The genome sequence of the food-borne pathogen Campylobacter jejuni reveals
 A:Reference number: A81250; MUID:20150912
 A:Accession: F81343
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-395 <PAR>
 A:Cross-references: GB:AL139076; GB:AL111168; NID:96968128; PIDN:CA872997.1; PID:669
 A:Experimental source: serotype O2, strain NCTC 11168
 C:Genetics:
 A:Gene: Cj0723c

Query Match 21.4%; Score 501.5; DB 2; Length 395;
 Best Local Similarity 31.2%; Pred. No. 6.7e-31;
 Matches 135; Conservative 88; Mismatches 154; Indels 55; Gaps 12;

QY 18 IISGFSIAQSFESYLTROYOKLSETKLPVLEDEIDETFKSRNYSRAKAFSIFGD 77
 DB 13 LLSWISYAOIRFLEREKDKOAILSEKDYONADIAIENE-----KFKLFSN 59

QY 78 VYNLAQKLVFKYDLPKIMHMAVSLNANLVPFRHNVSTVAOSLCFLGLLSSTLVYL 137
 DB 60 FYNLISIAWIGGF---LYLKEILLN-----NTRENPLFLSFLIINSILN 106

QY 138 PLSYSHVLEKFGFNKLTVOIMTDMIKSLTAVAGGPIILYFLKIPDKPPTDPLMY 197
 DB 107 PLSTYSEFIKKAHGSMTYKFLTKDMSLTLILGFLLYLTCYDFDFFG--FW 164

QY 198 IMVFLVVOILAM--TIIPVIMPENKFTPLEDELKKSIESLADRVGFPLDKIFVIDG 255
 DB 165 IAAFIACITIIINLITPILAPLFPNKKKIDDENLKLKISSLKQCGFSANGYVYIDA 224

QY 256 SKRSSHNAAYFTGLPFTSKRIYLFETLVNSNSTDEITAVLAHEIGHMOKNIIVNVISQ 315
 DB 225 SKRDKRLNAYFQGL-FKSKRVYLFPTLLKALNERELLAVLGHGHPVHKDILKALFNGA 283

QY 316 LHTFLISLFTSIYRNTSFYNTFGFLEKSTGSPVDPYITKEPFIIGFMFL--FNOLIT 372
 DB 284 ITMFLFFVFANL---PEF---VYLE---SHLEGNGGVFALL--FILANIFSPLIS 329

QY 373 PLECAMQVMSLISRTHEYQADAVAKKLYKONICRALIDLOKNIKSLTMVNDPLYSYH 432

Db 330 PM-----LNALSKRNEPAAQDHAKVTSKEDMKNALIALALAREKNAFKITKITEFPL 382
 Oy 433 SHPTLAERSTAL 444
 Db 383 SHPSISDRKAL 394

RESULT 7

Probable zinc-metallo proteinase - Helicobacter pylori (strain J99)
 C:Species: Helicobacter pylori
 A:Variety: strain J99
 C:date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 08-Oct-1999
 C:Accession: D71862
 R:Alm, R.A.; Ling, L.S.L.; Moir, D.T.; King, B.L.; Brown, E.D.; Doig, P.C.; Smith, D.R.; Ives, C.; Gibson, R.; Merberg, D.; Mills, S.D.; Jiang, Q.; Taylor, D.E.; Voyts, G.F.; Nature 397, 176-180, 1999
 A:Title: Genomic sequence comparison of two unrelated isolates of the human gastric path
 A:Reference number: A71800; MUID:99120557
 A:Accession: D71862
 A:Status: Preliminary
 A:Molecule type: DNA
 A:Residues: 1-407 <ARN>
 A:Cross-references: GB:AE001528; GB:AE001439; NID:g4155575; PIDN:AA06576.1; PID:g415558
 A:Experimental source: strain J99
 A:Genetics: Jhp0999

Query Match 19.6%; Score 459; DB 2; Length 407;
 Best Local Similarity 31.8%; Pred. No. 1.2e-27;

Matches 134; Conservative 81; Mismatches 148; Indels 58; Gaps 19;

Oy 39 OKLSETKLPVLEDEIDETFHKSRYNYSRAKAKFSIFGDVYVLAOKLVYIKYDLPRKIMW 98
 Db 35 OKLCE---KPVL---LPQDYEEAGNYAIRKMQLSIIQILD---GVIFAGM-VFEGLTN 84
 Oy 99 M--AVSLNAVLPVRFHMVSTVAQSLCFGLSLSTLVDPDLSYSHFVEEKGFKNL 156
 Db 85 LEDLTHYLN-LP---ETLGYLVFALLFLAIGSVLA---LPISYTTMHLDEKGFESKV 135
 Oy 157 TVQWLMTDMIKSLTLAAYAGPILYFLKIFDKPPTDFLWYIMV--LFVQOILAMTTP 214
 Db 136 SLSEFKDFPKGLLTLGVLITLITLMIIEVH--WEISSEFVVEVFMILANLFP 192
 Oy 215 VFMFMFKPTPLEDEGLKSLIESLADRYGFLDKIFVIDGSRSSHNAVYTGFLPFTSK 274
 Db 193 K-IAOLFQFTPLNNRDLSEQIESMMDKVGKSGIFVMDASKRQRLNAVYFGLG-KNK 250
 Oy 275 RVLFDTLVNSNSTDEITAVLAHEIGHMO-KNHYVMVYFSQHLTFLISLFTS---LY 329
 Db 251 RVLVDLTLISKVTEGLALAIIGHELGHFNKDLNKLGMGL-LALVFALLAHLPPLVF 309
 Oy 330 RMTSYNFGFLEKSTGSEFVDPVITKEPPIIGFM-LFNDLLTPLECAMQFVMSLSRTH 388
 Db 310 E-----GFNVSGTPASLI-----TILLFLPVESFVAMP-----IGFESK 346
 Oy 389 HEYQADAVAKKLYKONLCRALIDQIKNLSTMNVDPLVSSYVSHPTLAERSTALDYVS 448
 Db 347 NEYNADKFGASLSKRETLAKALVSIYENKAPYSHPFVFLHFTPLERLKLADYEI 406
 Oy 449 E 449
 Db 407 E 407

RESULT 8

F64567
 Zinc metalloproteinase (EC 3.4.24.-) - Helicobacter pylori (strain 26695)
 C:Species: Helicobacter pylori
 C:date: 09-Aug-1997 #sequence_revision 15-Aug-1997 #text_change 08-Oct-1999
 C:Accession: F64567

R:Tomb, J.F.; White, O.; Kerlavage, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann, Peterson, S.; Loftus, B.; Richardson, D.; Dodson, R.; Khaliq, H.G.; Glodek, A.; McK son, J.D.; Kelley, J.M.; Cotton, M.D.; Meldrum, J.M.; Fujii, C.; Bowman, C.; Watney A:Authors: Wallin, E.; Hayes, W.S.; Borodovsky, M.; Karpk, P.D.; Smith, H.O.; Fraser A:Title: The complete genome sequence of the gastric pathogen Helicobacter pylori.
 A:Reference number: A64520; MUID:97394467
 A:Accession: F64567
 A:Status: Preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-407 <TON>
 A:Cross-references: GB:AE000555; GB:AE000511; NID:g2313485; PIDN:AA07451.1; PID:g23
 C:Keywords: hydrolase; metalloproteinase

Query Match 19.4%; Score 454; DB 2; Length 407;
 Best Local Similarity 30.7%; Pred. No. 3e-27;

Matches 129; Conservative 84; Mismatches 151; Indels 56; Gaps 17;

Oy 39 OKLSETKLPVLEDEIDETFHKSRYNYSRAKAKFSIFGDVYVLAOKLVYIKYDLPRKIMW 98
 Db 35 OKLCE---KPVL---LPQDYEEAGNYAIRKMQLSIIQILD---GVIFAGM-VFEGLTN 84
 Oy 99 M--AVSLNAVLPVRFHMVSTVAQSLCFGLSLSTLVDPDLSYSHFVEEKGFKNL 156
 Db 85 LEDLTHYLN-LP---ETLGYLVFALLFLAIGSVLA---LPISYTTMHLDEKGFESKV 135
 Oy 157 TVQWLMTDMIKSLTLAAYAGPILYFLKIFDKPPTDFLWYIMV--LFVQOILAMTTP 214
 Db 136 SLSEFKDFPKGLLTLGVLITLITLMIIEVH--WEISSEFVVEVFMILANLFP 192
 Oy 215 VFMFMFKPTPLEDEGLKSLIESLADRYGFLDKIFVIDGSRSSHNAVYTGFLPFTSK 274
 Db 193 K-IAOLFQFTPLNNRDLSEQIESMMDKVGKSGIFVMDASKRQRLNAVYFGLG-KNK 250
 Oy 275 RVLFDTLVNSNSTDEITAVLAHEIGHMO-KNHYVMVYFSQHLTFLISLFTS---LY 330
 Db 251 RVLVDLTLISKVTEGLALAIIGHELGHFNKDLNKLGMGL-LALVFALLAHLPPLVF 310
 Oy 331 RMTSYNFGFLEKSTGSEFVDPVITKEPPIIGFM-LFNDLLTPLECAMQFVMSLSRTH 389
 Db 311 E-----GFNVSGTPASLI-----TILLFLPVESFVAMP-----IGFESK 347
 Oy 390 EYQADAVAKKLYKONLCRALIDQIKNLSTMNVDPLVSSYVSHPTLAERSTALDYSE 449
 Db 348 EYNADKFGASLSKRETLAKALVSIYENKAPYSHPFVFLHFTPLERLKLADYEI 407

RESULT 9

zinc metalloproteinase homolog [imported] - Halobacterium sp. NRC-1
 C:Species: Halobacterium sp. NRC-1
 C:date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
 C:Accession: E84192
 R:Ng, W.Y.; Kennedy, S.P.; Mahatiz, G.G.; Bergquist, B.; Pan, M.; Shukla, H.D.; Lask Jund, K.H.; Alam, M.; Freitas, T.
 Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000
 A:Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ehardt, H.; Lowe, T.M.
 A:Title: Genome sequence of Halobacterium species NRC-1.
 A:Reference number: A84160; MUID:20504483
 A:Accession: E84192
 A:Status: Preliminary
 A:Molecule type: DNA
 A:Residues: 1-428 <STO>
 A:Cross-references: GB:AE004437; NID:g10579955; PIDN:AA618905.1; GSPDB:GN00138
 C:Genetics: caax

Query Match 16.1%; Score 378; DB 2; Length 428;
 Best Local Similarity 29.3%; Pred. No. 2e-21;
 Matches 110; Conservative 64; Mismatches 143; Indels 58; Gaps 12;

OY 100 AVSLNLAIVLPREHNVSTVAOSICEL-----GLLSLSLTVLD-----LPLSYSH 144
 DB 68 SVYLAGAVLGLVY--TGAFGAANAFAVDVPGSDVLACVALLVAGTAVMAQVAPVPEDAVET 125
 OY 145 FVLEERKGFNKLTVQWITMIRKSLTLA-----VAIGPILYLFKIFDKFPTDLWYIM- 199
 DB 126 FGVSASAGFNPDSPALFARQALLSAGLAVFVAVLGAVALVAALAE-----WNVVAA 179
 OY 200 -----VLFVAVQILAMTIIIPVIMPFNKFTPLEDGELKKSIESLADRVGFPDLKIFVID 254
 DB 180 TGIAGVFLATQVL-----VRRVWVMPLEFDVPEDEGLRDAIEDVDFRAGACQVYVMN 235
 OY 255 GSKRSSHNAVFTGLPFTSRIVLFDLVNSNSTDETAVLAHIGMOKNHIYNNMIES 314
 DB 236 ASSRSQSNMFFTEFGFAT-KRVVLEFDTLIDMGETELQAVLAHELAHMKNGHIMQTGGAA 294
 OY 315 QLHTEFLFSLFTSYRNTSFYNTGPFLEKSTGSDVDPVITKEPPIIGFMLEFNDLITPL 374
 DB 295 TLQGVVLFVASRLDAGWLYGMG-----VPEQP-AAAGLLLAGVWVLOPL 338
 OY 375 ECAMQFVMSLISRTHETQADAYA-KKLGKONLCRALIDQIKLSTMNNDPLSYSHYS 433
 DB 339 SRLTAPLQNRMLANEREAFAVDVWGGEPFLADALGALTSQLNGNPFHPYEAHFYQ 398
 OY 434 HPTLAER---STALD 445
 DB 399 HPVPERIRYLTAD 413

RESULT 10

C69831
 Conserved hypothetical protein ynfN - Bacillus subtilis

C:Species: Bacillus subtilis
 C:Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 15-Oct-1999
 C:Accession: C69831; 140008
 R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berta
 C.: Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Ch
 A.: Ehlich, S.D.; Emerson, P.T.; Ehtian, K.D.; Errington, J.; Faderl, C.; Ferrari, E.
 Nature 390, 249-256, 1997
 A:Authors: Foulger, D.; Fritzt, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Gallazzi, A.; Gall
 lech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.
 Koester, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois
 A:Authors: lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maneel
 Y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelli
 Rieger, M.; Rivolta, C.; Kocha, E.; Roche, B.; Rose, M.; Sadale, Y.; Sato, T.; Scanlon
 A:Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seron
 akuch, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama
 T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K
 A:Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.
 A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.
 A:Reference number: A69580; M01D:98044033
 A:Accession: C69831
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-426 <KUN>
 A:Cross-references: GB:Z99109; GB:AL009126; NID:g26333260; PIDN:CAB12869.1; PID:el183031
 R:Smith, H.; de Jong, A.; Bron, S.; Venema, G.
 Gene 70, 351-361, 1988
 A:Title: Characterization of signal-sequence-coding regions selected from the Bacillus s
 A:Reference number: I39994; M01D:98108019
 A:Accession: I40008
 A:Status: preliminary; translated from GB/EMBL/DDBT
 A:Molecule type: DNA
 A:Residues: 1-42; 'LESTQAQ' <RES>
 A:Cross-references: GB:M22914; NID:g143697; PIDN:AAA22830.1; PID:g143698
 C:Genetics:
 A:Gene: ynfN

Query Match 14.0%; Score 328; DB 2; Length 426;
 Best Local Similarity 26.9%; Pred. No. 1.3e-17;

Matches 108; Conservative 76; Mismatches 152; Indels 66; Gaps 14;
 OY 81 LAQKLVFIKIDLF-----PKIMHNAVSL-----NAVLPREHNVSTVAOSLCF 124
 DB 53 VABOYNWKNKFLFEGVPLMFLFVLLVSGSKRIKMIKIAAPVPRF-----LQTVGF 106
 OY 125 LGLLSLSTLVLDPLSYSHFVLEERKGFNKLTVQWITMIRKSLTLAIGG---PILY 181
 DB 107 VFVSLTTLVTLPLDMIGVQVSLD-YNISQTTASNAKQVISEWISFPLTCLVLEY 165
 OY 182 LFLKIFDKFPTDLFWMYIMV-----FLFVAVQILAMTIIIPVIMPFNKFTPLEDGELKKSIE 237
 DB 166 WLKIRHEKMKMVLAMLTVPFSLFLEFIQ-----VPIIDPLXNDYPLKKNKESEKIL 218
 OY 238 SLADRVGFPDLKIFVIDGSKRSSHNAVFTGLPFTSRIVLFDLVNSNSTDETAVLAH 297
 DB 219 ELADENAPIADHYEVNMSEKTNALNAVYIGIG-ANKRIYLMPTLKLKLDSELTFTMGH 277
 OY 298 EIGHMOKNHIYNNVIFSQLHTFLFSL-----FTSYRNTSFYNTGPFLEKSTGSDVDP 352
 DB 278 EMGHYVKKH-----VYIGLAGYLLVSLAGFVIDKIKRYRLTRSFHLEGRDLAALP 332
 OY 353 VITKEPPIIGFMLEFNDLITPLECAMQFVMSLISRTHETQADAYAKKL-GYKQMLCRALI 411
 DB 333 L-----LLLLFVLSFAVMPFSNA-----VSRQENKADQYGIETLENREAAVKTFO 379
 OY 412 DLQIKNLSTMNNDPLSYSHYSKSHPTLAERSTALDYSEKKKN 453
 DB 380 DLAVTGLSOVDPVLYKIFRGSHPSIMERIOH---AEKEEN 417

RESULT 11

H75127
 CAAX prenyl proteinase 1 related PAB0555 - Pyrococcus abyssi (strain Orsay)

C:Species: Pyrococcus abyssi
 C:Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 07-Dec-1999
 C:Accession: H75127
 R:anonymous, Genoscope
 Submitted to the EMBL Data Library, July 1999
 A:Description: Pyrococcus abyssi genome sequence: Insights into archaeal chromosome s
 A:Reference number: A75001
 A:Accession: H75127
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-338 <RAW>
 A:Cross-references: GB:AJ248285; GB:AL096836; NID:g5458067; PIDN:CAB49737.1; PID:el51
 A:Experimental source: strain Orsay
 C:Genetics:
 A:Gene: PAB0555

Query Match 8.4%; Score 197; DB 2; Length 338;
 Best Local Similarity 22.3%; Pred. No. 1e-07;

Matches 88; Conservative 66; Mismatches 133; Indels 108; Gaps 18;
 OY 96 IMHNAVSLNAVLPVRF--HMVSTVA-----QSCLFLGLSLSLTVLPLSY 142
 DB 2 LMEWLI-LFGWILPVLFWHRHRIKELRGYKGRMIKEKVLFEALLISLVFFSLAGILGF 60
 OY 143 SHFLEERKGFNKLTVQWITMIRKSLTLAIGGPTL---VLFKIFDKFPTDLWYIM 197
 DB 61 FDFPMSTIEPLDKITVILPVLVPLVSTLLITRELGEKIRGDIYVAVF----- 110
 OY 198 IMVFLFVV---QLAMTIIIPVIMPFNKFTPLEDGELKKSIESL 239
 DB 111 -IVIFILFVLSLFTIPMPYPIVYLLMFLSPALFSLRVRFFGTGM-SGELAKVAEEL 168
 OY 240 ADRVGFPLDKIFVIDGSKRSSHNAVFTGLPFTSRIVLFDLVNSNSTDETAVLAHEI 299
 DB 169 CRRANVAVVEEYIID---EERIGAFVTGM--KCKTIFITKGALEKLNEMELLAVIAHEL 222
 OY 300 GHMOKNHIYNNVIFSQLHTFLFSLFTSYRNTSFYNTGPFLEKSTGSDVDPVITKEFP 359

A:Molecule type: DNA
A:Residues: 1-310 <ARN>
A:Cross-references: CB:AE001516; GB:AE001439; NID:g4155431; PIDN:AAD06444.1; PID:g4155431
A:Experimental source: strain J99
C:Genetics:
A:Gene: htpX
C:Superfamily: heat-shock protein htpX

Query Match 6.4%; Score 150.5; DB 2; Length 310;
Best Local Similarity 23.3%; Pred. No. 0.00033;
Matches 67; Conservative 57; Mismatches 106; Indels 57; Gaps 11.

```

Db      50 LMFQIFPT-----ITVMFVAIVILVCQNSSSITMLSGDEKLLDPISKVLSKKNOI 104
          |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
QY     241 DRVGFPL-----DKIFVIDSKRSSHNAVFTGLPTSKRIVLFDLVNSNDPEI 291
          |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
Db     105 HRLLELLEAKLFEFKLYIN---APYVMAFASGMDSSNLIATLSALIERLDEL 160
          |   |   |   |   |   |   |   |   |   |   |   |   |   |   |

```

[illegible]

RESULT 14
C69279
heat shock protein (hsp) homolog - *Archaeoglobus fulgidus*

C:\Species: Archaeoglobus fulgidus
C:\Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 20-Aug-1997
C:\Accession: 050370

R; Klelek, H. P.; Clayton, R. A.; Tomb, J. F.; White, O.; Nelson, K. E.; Ketchum, K. A.; Doña
C; Accession: U02715
R; Fleischmann, R. D.; Quackenbush, J. T.; Lee, N. H.; Sutton, G. G.; Gill, S.; Kirkness
R; Klelek, H. P.; Clayton, R. A.; Tomb, J. F.; White, O.; Nelson, K. E.; Ketchum, K. A.; Doña

Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weldman, J.F.; McDonald, L. Nature 390, 364-370, 1997

A.; Authors: Utteback, T.; Cotton, M.D.; Spriggs, T.; Artlach, P.; Kaine, B.P.; Sykes, Smith, H.O.; Woese, C.R.; Venter, J.C.

A;Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing arch
A;Reference number: A69250; M0ID:98049343

A:Accession: C69279
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
Molecular function: DNA

A; Residues: 1-335 <KLE>
A; molecule type: UNA
A; Cross-references: GR:AF001090. GR:AF0000782. NTD:c2689413. BTD:AA090998.1. BTD:c2655

C; Superfamily: heat-shock protein htpx

Query Match 6.48; Score 150; DB 2; Length 335;

Best Local Similarity 21.88; Pred. No. 0.0004;
Matches 71; Conservative 51; Mismatches 113; Indels 90; Gaps 13;

QY 163 TDMIKSLTAVAGSPILYLFK---IFDKPTDFLMYIMVFLFVVQILAMTIIIPVFLNP 219

Db 44 TSMVLGMLLAISAIYILILAVGVISF-----YGLIFLLINLMLYLSPIYINL 98

QY 220 MFNKTPLLEDGELKKSITESLADRVGFPPLDKIFVIDGSKRSSHNAVFYTGLEPFSKRIVLE 279

Db 99 SYG---AQRERLQWVNSVARRLNVKPPKAVV-----RSPNNAFAYGNFLTCKFVAVS 150

QY 280 DFLVNSSTDEITAVLAHEIGHMOKNHIVMVFISQLHTFLIFSLFTSIYRMTSEYNPBG 339
::: | ::| :||| | | : | :| :| :|

151 ESLMMLSQEELAVIGHEIGH--HKHRDNAV-----LLFGLLPSVI---FY--LG 195

QY 340 FFLEKST-----GSFVDPVITKEPPIIGFMLEFNDLITPLECANQFVMSLSISTRHEQYA 393

```

Db      196 YALHSSMRDRRCQAQLAIGIA--AVIVSFI-----VQILVAFSRLREXYA 241
QY      394 DAYAKKIGYKONLCRALIDL-----QI 415
Db      242 DFEQVRAITNKDAMORSIAKIHLYFHRYPDYLAPIODSKFRFLFYAFNNAVEPTTRADI 301
QY      416 KNLSTMNVDPLYSSYHSHPTLAER 440
Db      302 EALKMKVSPF-QEFLSTHPPPKR 325

```

RESULT 15

E81951

```

Probable membrane-bound zinc metalloproteinase (EC 3.4.24.-) NMA1031 [imported] - Neisser
C:Species: Neisseria meningitidis
C:Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 02-Feb-2001
C:Accession: E81951
R:Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Morel
; Holroyd, S.; Jørgensen, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajandream,
Nature 404, 502-506, 2000
A:Title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis Z2491.
A:Reference number: A81775; MUID:20222556
A:Accession: E81951
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-279 <PAR>
A:Cross-references: GB:AL162754; GB:AL157959; NID:97379424; PIDN:CAB84300.1; PID:9737973
A:Experimental source: serogroup A, strain Z2491
C:Genetics:
A:Gene: htpX; NMA1031
C:Superfamily: heat-shock protein htpX
C:Keywords: hydrolase; metalloproteinase

```

```

Query Match      6.3%; Score 148.5; DB 2; Length 279;
Best Local Similarity 25.7%; Pred. No. 0.00041;
Matches 59; Conservative 42; Mismatches 88; Indels 41; Gaps 10;

```

```

QY      228 EDGELKKSIESLADRVGFPLDKIEYIDGSKRSSHSNAYFTGLPTSKRIVLEDTLVNSNS 287
Db      75 EEAWLNTVEAQAQRMNKTPEVALY---HSPEPNAPATGASRNSSLIAVSTGLDHT 130
QY      288 TDEITAVIAHEIGHQKNIYNN--VIFSQHTFLFSLFTSYRNTSFYNTFGFLEKST 346
Db      131 RDEVEAVIAHEMAHNGNDMTLTLLIGVNTFVEF--LSRIANLIARNNDG--SOSQ 185
QY      347 GSFVDPVITKEFPITIGFMLFNDLTPLECAMQFWSLSRTHFYQADAYAKKILGYKONL 406
Db      186 GTYF--LVSMVFQILFGL-----ASLIYVMFSPQREYRADAGAAKILYGAPKM 231
QY      407 CRAL-----IDLQIKNLSTMNV-----DPLYSYHSHPTLAERSTAL 444
Db      232 ISALQRLKGNVVDLP--EEMNANMGIAGDTRDSILS---THPSLDNRIRAL 276

```

```

Search completed: February 12, 2002, 09:26:37
Job time: 63 sec

```

THIS PAGE BLANK (P. 370)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 12, 2002, 09:26:37 ; Search time 26.02 Seconds

(without alignments)
922.175 Million cell updates/sec

Title: US-09-165-460a-4

Perfect score: 1651

Sequence: 1 MLOFSTFLVLYTISIVLP.....ISLKDRLQTLVGNPGYRITL 315

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

PIR_68:*
1: PIR1:*
2: PIR2:*
3: PIR3:*
4: PIR4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|-------------|
| 1 | 1651 | 100.0 | 315 | 2 | SS4486 |
| 2 | 202.5 | 12.3 | 271 | 2 | T37745 |
| 3 | 113.5 | 6.9 | 339 | 2 | T19663 |
| 4 | 109 | 6.6 | 1363 | 2 | T15653 |
| 5 | 104.5 | 6.3 | 255 | 2 | B81693 |
| 6 | 102.5 | 6.2 | 1471 | 2 | T40117 |
| 7 | 100.5 | 6.1 | 433 | 2 | T11162 |
| 8 | 98 | 5.9 | 256 | 2 | H71537 |
| 9 | 97.5 | 5.9 | 433 | 2 | T17654 |
| 10 | 97.5 | 5.9 | 589 | 2 | D86243 |
| 11 | 97 | 5.9 | 483 | 2 | A81352 |
| 12 | 96.5 | 5.8 | 226 | 2 | T09804 |
| 13 | 96.5 | 5.8 | 534 | 1 | S73804 |
| 14 | 96 | 5.8 | 328 | 2 | E86837 |
| 15 | 96 | 5.8 | 426 | 2 | S58304 |
| 16 | 95 | 5.8 | 274 | 2 | C83709 |
| 17 | 95 | 5.8 | 500 | 2 | T16630 |
| 18 | 94.5 | 5.7 | 224 | 1 | PMF657 |
| 19 | 94.5 | 5.7 | 244 | 2 | E69787 |
| 20 | 94.5 | 5.7 | 269 | 2 | T22591 |
| 21 | 94 | 5.7 | 226 | 2 | T12002 |
| 22 | 94 | 5.7 | 346 | 2 | T28730 |
| 23 | 94 | 5.7 | 347 | 2 | D71690 |
| 24 | 94 | 5.7 | 400 | 2 | T25605 |
| 25 | 94 | 5.7 | 769 | 2 | S50966 |
| 26 | 93 | 5.6 | 645 | 2 | T11157 |
| 27 | 93 | 5.6 | 2376 | 2 | S48405 |
| 28 | 92.5 | 5.6 | 670 | 2 | T43784 |
| 29 | 92 | 5.6 | 715 | 2 | S10843 |

| | | | | | | |
|----|------|-----|------|---|--------|--------------------|
| 30 | 91.5 | 5.5 | 457 | 2 | T11283 | NADH dehydrogenase |
| 31 | 91.5 | 5.5 | 471 | 2 | A43956 | serotonin receptor |
| 32 | 90.5 | 5.5 | 492 | 2 | E58931 | NADH dehydrogenase |
| 33 | 90.5 | 5.5 | 852 | 2 | A72343 | hypothetical prote |
| 34 | 89.5 | 5.4 | 372 | 2 | A70302 | rod shape determin |
| 35 | 89.5 | 5.4 | 390 | 2 | T28083 | hypothetical prote |
| 36 | 89.5 | 5.4 | 449 | 2 | S02011 | serotonin receptor |
| 37 | 89.5 | 5.4 | 471 | 2 | A34863 | serotonin receptor |
| 38 | 89.5 | 5.4 | 471 | 2 | S40689 | 5-hydroxytryptam |
| 39 | 89.5 | 5.4 | 471 | 2 | S11280 | serotonin receptor |
| 40 | 89.5 | 5.4 | 638 | 2 | T11801 | NADH dehydrogenase |
| 41 | 89 | 5.4 | 347 | 2 | S47871 | NADH dehydrogenase |
| 42 | 89 | 5.4 | 385 | 2 | S43540 | YSD83 protein - ye |
| 43 | 89 | 5.4 | 2222 | 2 | A37490 | voltage-dependent |
| 44 | 88.5 | 5.4 | 312 | 2 | I33701 | olfactory receptor |
| 45 | 88.5 | 5.4 | 414 | 2 | T19912 | hypothetical prote |

ALIGNMENTS

RESULT 1

SS4486
probable membrane protein YMR274c - yeast (Saccharomyces cerevisiae)

N:Alternate names: hypothetical protein YMR156.16c

C:Species: Saccharomyces cerevisiae

C:Date: 08-Jul-1995 #sequence_revision 19-Oct-1995 #text_change 06-Feb-1998

C:Accession: S54486

R:Lyse, G.: Churcher, C.M.
submitted to the EMBL Data Library, May 1995

A:Accession number: S54014

A:Accession: S54486

A:Molecule type: DNA

A:Residues: 1-315 <LYE>

A:Cross-references: EMBL:249260; NID:9809081; PID:9809097; MIPS:YMR274C

A:Experimental source: strain AB972

C:Genetics:

A:Gene: SGD:RCCL

A:Cross-references: SGD:S0004887; MIPS:YMR274C

A:Map position: 13R

C:Keywords: transmembrane protein

F:7-23/Domain: transmembrane #status predicted <TM1>

F:45-61/Domain: transmembrane #status predicted <TM2>

F:106-122/Domain: transmembrane #status predicted <TM3>

F:210-226/Domain: transmembrane #status predicted <TM4>

F:241-257/Domain: transmembrane #status predicted <TM5>

F:280-296/Domain: transmembrane #status predicted <TM6>

Query Match 100.0% Score 1651; DB 2; Length 315;

Best Local Similarity 100.0% Pred. No. 2e-141; Mismatches 0; Indels 0; Gaps 0;

Matches 315; Conservative 0;

| | | | | |
|----|-----|----------------------|--|-----|
| QY | 1 | MLOFSTFLVLYTISIVLP | PLVATSOPESSKNDNPRTISRMOKLITIMLSMLVPLFQ | 60 |
| DB | 1 | MLOFSTFLVLYTISIVLP | PLVATSOPESSKNDNPRTISRMOKLITIMLSMLVPLFQ | 60 |
| QY | 61 | SQLSSTTSHISFKDAFLGII | PGYAAALPNMPOESQFVKDKLTKCAMLLTYCGPVLD | 120 |
| DB | 61 | SQLSSTTSHISFKDAFLGII | PGYAAALPNMPOESQFVKDKLTKCAMLLTYCGPVLD | 120 |
| QY | 121 | VLYHLNKRSSILEDYFHEFL | INWSPFRNIFAPRIETRYTSMLLTYNLTIPHSQLSQ | 180 |
| DB | 121 | VLYHLNKRSSILEDYFHEFL | INWSPFRNIFAPRIETRYTSMLLTYNLTIPHSQLSQ | 180 |
| QY | 181 | QLFMOPSLFPGLAHNAHNA | VEQLOEGSMPTVSLTLTTCFOILLTYTLFGGLTKFVPTGCGN | 240 |
| DB | 181 | QLFMOPSLFPGLAHNAHNA | VEQLOEGSMPTVSLTLTTCFOILLTYTLFGGLTKFVPTGCGN | 240 |
| QY | 241 | LMCCITLHALCNIMGFPGR | SRNLNHTVVDKRAGRISKLVSIMNKCYPALVLGLISLKD | 300 |
| DB | 241 | LMCCITLHALCNIMGFPGR | SRNLNHTVVDKRAGRISKLVSIMNKCYPALVLGLISLKD | 300 |

A: genome: mitochondrion
A: genetic code: SGCA
C: Superfamily: H⁺-transporting ATP synthase protein 6

RESULT 14
E86837

two-component system regulator l1rh [imported] - *Lactococcus lactis* subsp. *lactis* (strain C:\Species: *Lactococcus lactis* subsp. *lactis*
C:\Date: 23-Mar-2001 #sequence_revision 23-Mar-2001 #text_change 23-Mar-2001
C:\Accession: E86837
R:\Bolotin, A.; Winkler, P.; Mauger, S.; Jallion, O.; Malarme, K.; Weissensbach, J.; Ehrlich
Genome Res. In press, 2001
A:\Title: The complete genome sequence of the lactic acid bacterium.
A:\Reference number: AB6625
A:\Accession: E86837
A:\Status: preliminary
A:\Molecule type: DNA
A:\Residues: 1-328 <STO>
A:\Cross-references: GB:AE005176; NID:912724717; PIDN:AAK05799.1; GSPDB:GN00146
A:\Experimental source: strain IL1403
C:\Genetics:
A:\Gene: l1rh

| | | | | |
|-----------------------|------------------|----------------|------------|-------------|
| Query Match | 5.8%; | Score 96; | DB 2; | Length 328; |
| Best Local Similarity | 25.5%; | Pred No. 0.4; | | |
| Matches 53; | Conservative 31; | Mismatches 66; | Indels 58; | Gaps 12; |

```
QY      4 FSTFLVLLY-ISISVLPLATSOPEGSKRNPRTISRMOKITIMLINLFLVFPLOQ 62
       ::|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db      7 YTTFLTILYTIVMSLSLSLY-----LKER--ENNYLLVSYPFLILLINGL 49
```

```

Oy      63 L-SSTSHI---SFKDAFLGLGIIPG-YAALBNPWOFSQFVKDITKCVAMLLTV-CG 115
      : | | : | | | | : | : | |
Db      50 IVTMSETFQIFANDYNOGFVNPDIETIYYLA-----TFGALKLVTELFEEKITTYOYG 104

```

```

Oy  116 PVLDEVLYLHLNPKSSILEDYHEFLNIMWSFRNFIFAPITEEIFYNSMLLTYYL----- 169
      : | : : : | : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db  105 IYIVFALMLIYVPEMA-----NSALKWLYL-----YFSSQLTFYVLGTYLLI 146

```

```

Oy 170 ---NLPHSQL-SYQQLFWQPSLFFGLA 193
      :| | | | :| :| | | |
Db 147 HNRKMIPKSSLGKYVKLIGSLAIFFGLA 174

```

RESULT 15

hypothetical protein SPAC18b11.05 - fission yeast (Schizosaccharomyces pombe)
C:Species: Schizosaccharomyces pombe
C:Date: 14-Jan-1996 #sequence_revision 01-Mar-1996 #text_change 21-Jan-2000
C:Accession: J37909; 558304
R:Devlin, K.; Churcher, C.M.; Barrell, B.G.; Rajandream, M.A.; Walsby, S.V.
submitted to the EMBL Data Library, August 1995
A:Reference number: Z21753
A:Accession: J37909
A:Status: preliminary; translated from GR/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-426 <DE2>
A:Cross-references: EMBL:Z50728; NID:9929886; PID:CAA90590.1; PID:9929891; GSPDE:GNO006
A:Experimental source: strain 972h-; cosmid c18b11
A:Genetics:
A:Gene: SPAC18b11.05
A:Map position: 1L

| | | | | |
|-----------------------|--------|-----------------|-------|----------------|
| Query Match | 5.8%; | Score 96; | DB 2; | Length 426; |
| Best Local Similarity | 21.9%; | Pred. No. 0.54; | | |
| Matches | 58; | Conservative | 40; | Mismatches 93; |
| | | | | Indels 74; |
| | | | | Gaps 12; |

```

QY      11  LYSISVLPVATSQPEGSKRONPRTIKSRMOKLTIMLISNLELVPPLOSQLSSTTSHI  70
      :|:|:|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db      179  IYMSGYTESLEPAFSLG-----LLFIKKQOYPAAFLWSLATLIRSNQ  22

```

```

QY      71 SFKAEFLGIIIPGYAALPNP-----WQFSQFVKDLTKCAMLTLVCGPVLDFVLY--- 123
      - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
Db      224 IFWCIFFGMAIGTLKISLERQLTPEOVQSOLVGYTKCLILVPEFYNYQLGFKFLCPG 283

```

```
QY      124 -----HLLNPKSSILEDFFYH-EPLNIWSFR--NFIFA-----PITEEIFY--TSML 164
```

```

Db      284  VAMCKRSFLIYP--AVGEKXWNNGLRWYPLNNIPNLFALLSIIPILFALFESIGST 344
QY      165  LTTYYNLPHQSLOVQOLFQNPSPFEGALNHNHAEVDQESMPTYSILLTTCQIIYTT 22
      ||:::|||||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db      342  LHSFSSIKSHLVLS--ALY-----LYIGCFHH-----TQVLRMSALPILYWS 384
QY      225  L-----FGLTFKEFVY 235
      ||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db      385  MAHATLYAKSRNLKAFGHCILFEVHI 409

```

Search completed: February 12, 2002, 09:26:40
Job time: 66 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 12, 2002, 13:20:11 : Search time 4622.84 seconds

(without alignments)
10520.305 Million cell updates/sec

Title: US-09-165-460a-3

Perfect score: 2948
Sequence: 1 tgaactgttgatgaacaaag.....gggagagataaagaatcaca 2948

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Maximum Match 0%
Listing first 45 summaries

Database :

GenBdb1:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_da:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_om:*
20: em_or:*
21: em_ov:*
22: em_pat:*
23: em_ph:*
24: em_pl:*
25: em_ro:*
26: em_sts:*
27: em_sy:*
28: em_un:*
29: em_vi:*
30: em_htgo_hum:*
31: em_htgo_inv:*
32: em_htgo_rod:*
33: em_htg_hum:*
34: em_htg_inv:*
35: em_htg_rod:*
36: em_htg_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB | ID | Description |
|------------|--------|-------------|--------|----|-------------|--------------------|
| 1 | 2948 | 100.0 | 29987 | 8 | SC8156 | 249260 S.cerevisia |
| 2 | 1560.4 | 52.9 | 4156 | 8 | D50276 | D50276 Saccharomyc |
| 3 | 1361 | 46.2 | 4130 | 8 | YSCNCRIG | L42821 Saccharomyc |
| 4 | 1163 | 39.5 | 4225 | 8 | SCU65682 | U65682 Saccharomyc |
| 5 | 1069.4 | 36.3 | 3250 | 8 | SCU32580 | U32580 Saccharomyc |
| 6 | 931 | 31.6 | 3525 | 8 | SCU63849 | U63849 Saccharomyc |
| 7 | 786.8 | 26.7 | 3240 | 8 | YSCBUL1 | D50083 Saccharomyc |
| 8 | 703.8 | 23.9 | 3020 | 8 | SCRD51GN | X88901 Yeast BUL1 |
| 9 | 225.2 | 7.6 | 270 | 8 | YSCGNC8X | M87379 Yeast Eco R |
| 10 | 143.8 | 4.9 | 27559 | 8 | SCYNR06C | Z71663 S.cerevisia |
| 11 | 75 | 2.5 | 2023 | 8 | SC8339 | 249210 S.cerevisia |
| 12 | 71.8 | 2.4 | 4775 | 8 | SCU32938 | U32938 Saccharomyc |
| 13 | 68 | 2.3 | 780 | 11 | CNS06KRB | AL402561 7' end of |
| 14 | 65.6 | 2.2 | 256172 | 2 | AC005139 | AC005139 Plasmodiu |
| 15 | 65.6 | 2.2 | 310779 | 2 | AC005140 | AC005140 Plasmodiu |
| 16 | 64.4 | 2.2 | 904 | 11 | CNS06189 | AL399667 73 end of |
| 17 | 64.4 | 2.2 | 86827 | 3 | PFMAL3P5 | AL034556 Plasmodiu |
| 18 | 63.4 | 2.2 | 1920 | 3 | PFGRUB | X62393 P.falciparu |
| 19 | 63.4 | 2.2 | 110000 | 3 | PFMAL13P2_0 | AL049185 Plasmodiu |
| 20 | 62 | 2.1 | 253305 | 2 | PFMAL3P7 | AL034559 Plasmodiu |
| 21 | 61.6 | 2.1 | 1141 | 6 | AX083744 | AX083744 Sequence |
| 22 | 60.8 | 2.1 | 234112 | 3 | PFMAL4P2 | AL035475 Plasmodiu |
| 23 | 60.6 | 2.1 | 202872 | 2 | AC016160 | AC016160 Homo sapi |
| 24 | 60 | 2.0 | 5632 | 3 | AF310892 | AF310892 Dictyoste |
| 25 | 59.4 | 2.0 | 3214 | 3 | AF019980 | AF019980 Dictyoste |
| 26 | 59.4 | 2.0 | 318221 | 2 | PFMAL13P3 | AL034576 Plasmodiu |
| 27 | 58.8 | 2.0 | 321003 | 2 | PFMAL4P3 | AL034576 Plasmodiu |
| 28 | 58.4 | 2.0 | 78439 | 17 | HSUB5197 | U85197 Homo sapien |
| 29 | 58.4 | 2.0 | 169546 | 2 | AC004157 | AC004157 Plasmodiu |
| 30 | 58.4 | 2.0 | 245802 | 2 | AC006279 | AC006279 Plasmodiu |
| 31 | 58.4 | 2.0 | 251124 | 3 | HUAE000660 | AE000660 Homo sapi |
| 32 | 58.2 | 2.0 | 153098 | 3 | PFMAL3P2 | AL034558 Plasmodiu |
| 33 | 58 | 2.0 | 169794 | 2 | AC004688 | AC004688 Plasmodiu |
| 34 | 58 | 2.0 | 178137 | 9 | AC011302 | AC011302 Homo sapi |
| 35 | 57.8 | 2.0 | 13684 | 3 | AE001403 | AE001403 Plasmodiu |
| 36 | 57.8 | 2.0 | 153477 | 2 | AC006278 | AC006278 Plasmodiu |
| 37 | 57.6 | 2.0 | 53932 | 2 | AC023371 | AC023371 Homo sapi |
| 38 | 57 | 1.9 | 12029 | 3 | AE001431 | AE001431 Plasmodiu |
| 39 | 57 | 1.9 | 163443 | 2 | AC006280 | AC006280 Plasmodiu |
| 40 | 57 | 1.9 | 178273 | 2 | AC005308 | AC005308 Plasmodiu |
| 41 | 57 | 1.9 | 196149 | 2 | AC004709 | AC004709 Plasmodiu |
| 42 | 56.8 | 1.9 | 67970 | 3 | PFMAL1P3 | AL031746 Plasmodiu |
| 43 | 56.8 | 1.9 | 149627 | 2 | AC087428 | AC087428 Homo sapi |
| 44 | 56.6 | 1.9 | 92633 | 2 | PFMAL4P1_3 | Continuation (4 of |
| 45 | 56.6 | 1.9 | 153477 | 2 | AC006278 | AC006278 Plasmodiu |

ALIGNMENTS

| RESULT | 1 |
|------------|--|
| SC8156/c | |
| LOCUS | SC8156 29987 bp DNA 11-AUG-1997 |
| DEFINITION | S.cerevisiae chromosome XIII cosmid 8156. |
| ACCESSION | Z49260 Z71257 |
| VERSION | Z49260.1 GI:809081 |
| KEYWORDS | BUL1; cytochrome b5; DAG1; initiation factor 1A; inorganic pyrophosphatase; IPP2; nitrate reductase; orotate phosphoribosyltransferase; tau element; TIF11; TRS3; transfer RNA-Gln; trehalose-phosphate synthase; U6 snRNP; URA10. |
| SOURCE | baker's yeast. |
| ORGANISM | Saccharomyces cerevisiae |
| REFERENCE | 1 (bases 1 to 29987) |
| AUTHORS | Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; Saccharomycetaceae; Saccharomycetaceae; Saccharomycetes. |
| JOURNAL | Unpublished |
| REFERENCE | 2 (bases 1 to 29987) |
| AUTHORS | Barrell, B. and Rajandream, M.A. |
| TITLE | Direct Submission |

JOURNAL

COMMENT

Submitted (12-MAY-1995) *Saccharomyces cerevisiae* chromosome XIII sequencing project, Sanger Centre, Hinxton Hall, Hinxton, Cambridge CB10 1RQ. E-mail: barrell@sanger.ac.uk

Notes:

All CDS over 100 codons have been analysed. CDS that are completely overlapped and those that are overlapped by more than 50% of their length by a larger CDS have been omitted from this analysis.

Details of the omitted CDS are available on request. The more significant matches with motifs in the PROSITE database are also included but some of these may be fortuitous. The length in codons and the calculated codon adaptation index (CAI) is given for each CDS.

Cosmid 8156 is overlapped at the start by cosmid 9920, emb1 entry SC9920,

accession no. 248639 and at the end by cosmid 8021.

FEATURES

source

Location/Qualifiers

1. 29987

/organism="Saccharomyces cerevisiae"

/strain="AB972"

/db_xref="taxon:4932"

/chromosome="XIII"

/clone="cosmid 8156"

/map="13R"

complement(<1..356)

/note="YH8156.01c, unknown, len: 118, CAI: 0.14, overlaps and extends YH9920.13c"

/codon_start=1

/product="unknown"

/protein_id="CA89242.1"

/db_xref="GI:809082"

/db_xref="SWISS-PROT:Q03496"

/translation="MTDVOFSSQELIFRVKEFLIANPKNINENKDAVLTOIEDH
FRLIYIKDGLNLESTRLIFPDPSICLLSHQIASKKIDISQEFSLAVKEQLLT
EANANIFEEVLDLFWMA"

1. 140

/note="overlap with SC9920 248639 S.cerevisiae chromosome XIII cosmid 9920"

complement(851..1312)

/gene="TIF1"

complement(851..1312)

/gene="TIF1"

/note="YH8156.02c TIF1 gene, IFLA-YEAST, P3891, initiation factor 1a"

/codon_start=1

/product="Tif1p"

/protein_id="CA89243.1"

/db_xref="GI:809083"

/db_xref="SWISS-PROT:P38912"

/translation="MGKNTGKGRGKRGKNDSDPKRELIVKEGQEVATITKMLGN
GVASCFDGNKMAHIRGKLRKVMGGDILVSLRDFODDCDVHRYNDEART
LKNQGLPENAKINETDNFGFSEDEVDNFEFGNADDEDEGDELDIDDI"

complement(1678..4842)

/gene="TPS3"

complement(1678..4842)

/gene="TPS3"

/note="YH8156.03c, TPS3 gene, len: 1054, CAI: 0.21, SW:TPS3_YEAST_P38426, alpha, alpha-trehalose-phosphate synthase and PIR:S27471, P800213 Lipocalin signature; translation in TPS3_YEAST starts at amino acid 33 in this sequence"

/codon_start=1

/product="TPS3p"

/protein_id="CA89244.1"

/db_xref="GI:809084"

/db_xref="SWISS-PROT:P38426"

/translation="MTIIVASLFLPYTPPEADVTNSDTAKIVSSMIKVCNNQELS
NKKQSSSVTSASHYIGLPQAOINGEPLORANVSPATGVYNNMEMLSSEDFL
EELTANATHAANGGIPANNPVSSGSTAOPRVEFFESAPARVCSPOASASSTISA
SRSAHHNLSLSLKKRIVSGTADGSSIASSSNLYRSCQFODNFIEDTDEDDIDSLE
REFFOOASLPSKRVSGTADGSSIASSSNLYRSCQFODNFIEDTDEDDIDSLE
TDATKRYNVPKFGGYNNAKLRASLNRNSYELFKHLPTIYDSKNGSLKNNVNIAY
AEKTVKEPVSWGTGICPTDELPHVEVCHKISKLEDDFSSFPVVTDDITFKGAYKNYA

misc-feature

CDS

/note="P500213 Lipocalin signature"

5199..6140

/note="YH8156.04, unknown, len: 313, CAI: 0.11, previously noted ORF, SW:YTP3_YEAST_P38430"

/codon_start=1

/product="unknown"

/protein_id="CA89245.1"

/db_xref="GI:809085"

/db_xref="SWISS-PROT:P38430"

/translation="MNLVDHCHVITDPDNTTCGGDGSQGLRCVSSNPYDWNLT
KLAGRSTKNDICVGFVHPVWSHLEYSRDVSHYODVEYKNEEFDLSVYLT
PEPLDEEYIKRENDTLVIGEIGDLKFLRPLANFVQNEKARLTITVYKLSHOB
TVPRFRRLARLHSPISIHVDHCHGKLNICBELLITVYSKICHSYSGKSTLLG
QMLKPPDRIFVSLSKINFPKPEBDAVRSIPSCILITENDYPIDNPDSYQKAL
TEOLOYINAOIARAWDETLDASQALAVYENFQKFK"

6392..6997

/note="YH8156.05, unknown, len: 301, CAI: 0.12, SW:YTP2_YEAST_P38429 previously noted ORF"

/codon_start=1

/product="unknown"

/protein_id="CA89246.1"

/db_xref="GI:809086"

/db_xref="SWISS-PROT:P38429"

/translation="MARPVNTNAETESRGRPTGGGAYANNGSCNNNGNNNNNN
NNNNNNNNNNNNNNPSSSGRTNGKOTLAQOQYIKNLETHITDHPDLRKSNN
DPEEYDARLRKRDHQLDVPNNLTQGLVLSKLGAKRTYSTRNTQGHDRIRKR
DLANVVRHFDENSIKTDCTIPQITKVKQKKRFPKFK"

7278..7889

/note="YH8156.06, unknown, len: 301, CAI: 0.13, SW:YTP1_YEAST_P38428, extends previously noted ORF"

/codon_start=1

/product="unknown"

/protein_id="CA89247.1"

/db_xref="GI:809087"

/db_xref="SWISS-PROT:P38428"

/translation="MEDSRLLITLILVGVIFLKKFPFOSNHPASAORLSTGVANHR
PGSTQNALRRTGNGVGHVTVQWETVONLAPNLHPDRISLDEGTVEETERY
LRGDEFSFPGFEPSPRPMKANAVDNNAGGGEFDPKRNKNICENLIDKHFVDLN
EDMSNLSFKDLIDERRKLLVQARRNLEKIOSDLOSLL"

complement(8013..9398)

/note="YH8156.07c, unknown, len: 461, CAI: 0.11"

/codon_start=1

/product="unknown"

/protein_id="CA89248.1"

/db_xref="GI:809088"

/db_xref="SWISS-PROT:Q03508"

/translation="MEFEFRRKGESSCGNHRVLRKKNSCACELAVFEVEBAH
TKLHIDFQNTLFTAGPTPOLYTRRELINLTSTLPNGGMNPEGLAIEISKTK
PRYSNMADIVKLAESYSKADITISVILVIGRESKTHKLEHALLQTRSHKSENDE
RFNAVCLKRAISLEYSTKTKELMRDFLEYPSLRLELSYDDRIHOIDAKSFHSLD
LPLKMSAIPVPRPTSLPREDELVMVMVARNNSQALSTQKPDLRPRQIGVIL
CFASHRLSLIEVYLYLKRGRGRTFRPKIYEHLVYICAPGPGDIPALELAKWSSND
TRTFDEKQVOHISOLEYLEOPKCIHVQOVPLAYIASHHNRKPLVEYKATPEP
NRYTDLPEFYITGTHFYKRDRIEDLEVYTERLINCKEDIHWPPLDNTPIKAFERF
AKAALPCNSA"

9991..12852

/note="YH8156.08, unknown, len: 953, CAI: 0.20, probable integral membrane glycoprotein"

/codon_start=1

/product="unknown"

/protein_id="CA89249.1"

/db_xref="GI:809089"

/db_xref="SWISS-PROT:Q03508"

/translation="MEFEFRRKGESSCGNHRVLRKKNSCACELAVFEVEBAH
TKLHIDFQNTLFTAGPTPOLYTRRELINLTSTLPNGGMNPEGLAIEISKTK
PRYSNMADIVKLAESYSKADITISVILVIGRESKTHKLEHALLQTRSHKSENDE
RFNAVCLKRAISLEYSTKTKELMRDFLEYPSLRLELSYDDRIHOIDAKSFHSLD
LPLKMSAIPVPRPTSLPREDELVMVMVARNNSQALSTQKPDLRPRQIGVIL
CFASHRLSLIEVYLYLKRGRGRTFRPKIYEHLVYICAPGPGDIPALELAKWSSND
TRTFDEKQVOHISOLEYLEOPKCIHVQOVPLAYIASHHNRKPLVEYKATPEP
NRYTDLPEFYITGTHFYKRDRIEDLEVYTERLINCKEDIHWPPLDNTPIKAFERF
AKAALPCNSA"

9991..12852

/note="YH8156.08, unknown, len: 953, CAI: 0.20, probable integral membrane glycoprotein"

/codon_start=1

/product="unknown"

/protein_id="CA89249.1"

/db_xref="GI:809089"

/db_xref="SWISS-PROT:Q03508"

/translation="MEFEFRRKGESSCGNHRVLRKKNSCACELAVFEVEBAH
TKLHIDFQNTLFTAGPTPOLYTRRELINLTSTLPNGGMNPEGLAIEISKTK
PRYSNMADIVKLAESYSKADITISVILVIGRESKTHKLEHALLQTRSHKSENDE
RFNAVCLKRAISLEYSTKTKELMRDFLEYPSLRLELSYDDRIHOIDAKSFHSLD
LPLKMSAIPVPRPTSLPREDELVMVMVARNNSQALSTQKPDLRPRQIGVIL
CFASHRLSLIEVYLYLKRGRGRTFRPKIYEHLVYICAPGPGDIPALELAKWSSND
TRTFDEKQVOHISOLEYLEOPKCIHVQOVPLAYIASHHNRKPLVEYKATPEP
NRYTDLPEFYITGTHFYKRDRIEDLEVYTERLINCKEDIHWPPLDNTPIKAFERF
AKAALPCNSA"

[illegible]

| | | | | |
|----------------------------|---------|--------------|-------|---------------|
| Query Match | 100.0%; | Score 2948; | DB 8; | Length 29987; |
| Best Local Similarity | 100.0%; | Pred. No. 0; | | |
| Matches 2948; Conservative | 0; | Mismatches | 0; | Gaps 0; |

| | | | |
|----|-------|--|------|
| QY | 481 | caaacgagacataatgctgtaaccttggaattcaataagagacataaggaaccttaattcc | 540 |
| Db | 27305 | CAAAACGAGACATTATGTGAACCTTGGAAATTTATTAAGGACATTAAAGAACTTTAATTC | 2724 |
| QY | 541 | aagtttgaagcgcgcctatgtgtgagttttatgtgcgttcgagaaatgaatlaatttga | 600 |
| Db | 27245 | AAGTTTGAAGCGCCTATGTTGTGAGTTTATTTATTCGCTCCAGTAATGAATTAATTTGA | 2718 |
| QY | 601 | aaacaccttbgcgtaagcagatactatcccatcttcgttaagcaagtgcacaaata | 660 |
| Db | 27185 | AAACACCTTGGCGTACCGCAAGATGTGATATCCCTATTTCGTTAGCGAAGTGACAAATA | 2712 |
| QY | 661 | aaaaacattagaaaaaaatctcgttacttttcttataagatalagatalgtatggttc | 720 |
| Db | 27125 | AAAAACATTAGAAAAAATCTCGTTACCTTTCTTATAGATATATGATATATGTATGTTT | 2706 |
| QY | 721 | gcttatgatagaaggtatttatcgcgctctcttgatctccctatatataaatctcttc | 780 |
| Db | 27065 | GCTTATGATGATGAAGGTATTTATGCGCTCCCTGTTATTCCTATTTTATTAATTAATTCGTT | 2700 |
| QY | 781 | taaatgcattctcttgctgctctctcttgctgcctgcgtatatcttctcttttttggaacag | 840 |
| Db | 27005 | TAAATGCAATTTCTGSGTGCCTTTGTTGTCCTTCGTATATTTTTTTTTTTTGACACACTG | 2654 |
| QY | 841 | gattgaaaaaccttgatctatcttattacatttataagttactacaaatactgagatt | 900 |
| Db | 26945 | GATGGAATAACCTTGTGATGATTTTATTACCTTTATTTTAAAGTTACTAATAATATCGAATTT | 2688 |
| QY | 901 | caggaaacaaacatagaaattctctctgtcagaagaaaaataaaacgaataatctgagtct | 960 |
| Db | 26885 | CAGGAACAAAACATAGAAATTTCTTTGCTACAGAAAAATTAACGAAATTAATTTGATGCTT | 2682 |
| QY | 961 | tgactactgactctgcctcctagagagaacacgaagaagaaatgycacaaatctcaaatc | 1020 |
| Db | 26825 | TGACTACTGACTCTGCTTCTCATTAAGAGAAACGAAACGACAACTGCTACATTTCTAACATT | 2676 |
| QY | 1021 | ctcagtgctccctacatactccatactcctatctgtcaccgctatatactgaacttcacaac | 1080 |
| Db | 26765 | TCTATGTGCTCTATACATCTCCATATTCGATATGTCGATACGCGTATATGCACTTCACAAC | 2670 |
| QY | 1081 | agaagaggtctaaacgagataatccctcgaaacgataaactctgcattgcagaaaccttaacat | 1140 |
| Db | 26705 | AGAAGGGCTTAAGAAAGAAATATCTCTGAACAGATTAATCTCCCAATGCAAAAACTTACAAAT | 2664 |
| QY | 1141 | tatgcataatctcaaaccttttcttgctgccttttacaactctgaatctatctagtaaac | 1200 |
| Db | 26645 | TATGCTATTTTCCAACTTTTTTTTGGTGCTTTTTTACATCTCAATTTATCTAGTACAC | 2658 |
| QY | 1201 | ttcacataatgattcagaagacgcatlcttagccttagtactaccccaagttlaacgc | 1260 |
| Db | 26585 | TTCACATATATAGTTTCAAGAGACCATTTTATAGCGTTAGGTATATTCAGAGTATATAGCG | 2652 |
| QY | 1261 | tgcatctgcaaaacctctggcgaattcagccaagtctgtgaaagacttaacgaaaatgctgttc | 1320 |
| Db | 26525 | TGCATTCGCAAAACCCCTTGGCAATTCAGCAAGTTCGGAAAGACTTAAAGAAATGTGTTC | 2646 |
| QY | 1321 | gagcttatgacctatatactgtgagaccgcttttagactttgtatatactcatattaa | 1380 |
| Db | 26465 | GATGTTATTGACCTTATATTTGTGAGCCGTTTTAGATTTTGTATATATATCTATTATTTAA | 2640 |
| QY | 1381 | ttcaaaagagcctataactcttgaagatttttacaatgtaatcctgaaatactcttgagcttcag | 1440 |
| Db | 26405 | TCCAAAGAGCTCTATCTTGAAGATTTTTCACATGTAATTCGTGAATATTTGAGAGTTTCAG | 2634 |
| QY | 1441 | gaatttatatattgacccaataaacctgagagaataatttacaacgtcaatgcttctgaactac | 1500 |
| Db | 26345 | GAAATTTATATTTTCACCAATAACTCAGGAATAATTTTACAGCTCAAAAGCTTTTGACTAC | 2628 |
| QY | 1501 | gtacttaaaccttaacgcgacttcggaacctaagactcaaacgattatttttggaacatc | 1560 |
| Db | 26285 | GTACTTAACCTTATACCGCATTTCCGAACTTAAGCTATCAACAGTTATTTTGGCAACCATC | 2622 |
| QY | 1561 | gcttttttttgactctgcagcagacacacatgcttatagacatataagacgttacgaagagctcat | 1620 |

| Db | 25145 | CCGCTAGAGACCTTCAGATGGGAGGGGTACACTCAGGAGTACAGATCTCCATTAATGACTAA | 25086 |
|------------|---|--|-------------|
| QY | 2701 | caaggtaactaactaggtatataaaaaaagggtgttcactcttcacatccaggaaatlaagtct | 2760 |
| Db | 25085 | CAAGGTACTACTTATAGGTATATAAAAAAGGTGTTCACCTCCATCCAGGAAATTTAATGCT | 25026 |
| QY | 2761 | aacgtataaagaaaaaacttlatatgggttctcccgcaatcaaacccctaaagtttaagcct | 2820 |
| Db | 25025 | AACGTATTAAAGAAAACCTTTATTATGGGTCCCGCAATCAACACCTTAACGTTTAAGCT | 24966 |
| QY | 2821 | gataattccctagaagctctgtacaagataacttacaanaatcaacaactaagcagacaatggt | 2880 |
| Db | 24965 | GATATTATTCCTAGACACTGTACACAGATCTTTACAAAATATACAACTAAGCAGACATGCT | 24906 |
| QY | 2881 | gaagataatgta1g9gaaatagcaatgtaaaaaataacgatat1ggaataatg9gagagataaa | 2940 |
| Db | 24905 | GAAGTATATGATGGGAATAGCAATGAATAAATAACGATATTGAGATTAATGGGAGAGATMAA | 24846 |
| QY | 2941 | gaatcaca 2948 | |
| Db | 24845 | GAATCACA 24838 | |
| RESULT | 2 | | |
| D50276 | | | |
| LOCUS | D50276 | 4156 bp | DNA |
| DEFINITION | Saccharomyces cerevisiae HST1 gene, complete cds. | PLN | 10-FEB-1999 |
| ACCESSION | D50276 | | |
| VERSION | D50276.1 | GI:1777313 | |
| KEYWORDS | HST1; high-copy suppressor of TFP sensitivity. | | |
| SOURCE | Saccharomyces cerevisiae (strain:RAY-3a) DNA. | | |
| ORGANISM | Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; Saccharomycetales; Saccharomycetaceae; Saccharomyces. | | |
| REFERENCE | 1 (bases 1 to 4156) | | |
| AUTHORS | Tsuchiya,E. | | |
| TITLE | Direct Submission | | |
| JOURNAL | Submitted (13-APR-1995) to the DDBJ/EMBL/Genbank databases. Elko Tsuchiya, Hiroshima University Faculty of Engineering, Department of Fermentation Technology; Kagami-Yama, Higashi-Hiroshima, Hiroshima 739, Japan (E-mail:etsuchi@ipc.hiroshima-u.ac.jp, Tel:0824-24-7868, Fax:0824-24-7868). | | |
| REFERENCE | 2 (bases 1 to 4156) | | |
| AUTHORS | Tsuchiya,E., Matsuzaki,G., Tsukao,A. and Miyakawa,T. | | |
| TITLE | Molecular cloning and characterization of S. cerevisiae HST1 gene | | |
| JOURNAL | Unpublished (1995) | | |
| REFERENCE | 3 (sites) | | |
| AUTHORS | Tsuchiya,E., Matsuzaki,G., Kurano,K., Fukuchi,T., Tsukao,A. and Miyakawa,T. | | |
| TITLE | The Saccharomyces cerevisiae SPS1 gene is involved in the tolerance to high concentration of Ca2+ with the participation of HST1/NRC1/BPR1 | | |
| JOURNAL | Gene 176 (1-2), 35-38 (1996) | | |
| MEDLINE | 97075905 | | |
| FEATURES | | | |
| source | Location/Qualifiers | | |
| | 1..4156 | | |
| | /organism="Saccharomyces cerevisiae" | | |
| | /strain="RAY-3a" | | |
| | /db_xref="taxon:4932" | | |
| | 946..3693 | | |
| | /gene="HST1" | | |
| | 946..3693 | | |
| | /gene="HST1" | | |
| | /standard_name="high-copy suppressor of TFP sensitivity" | | |
| | /codon_start=1 | | |
| | /protein_id="BAA08819.1" | | |
| | /db_xref="GI:1816450" | | |
| | /translation="MSNRDNEMLRTTSSDKAIASORDKRSEVLIAAOSLDNEIRSV | | |
| | KNKIRKTSIGMDLLIPDLIDKRGSGSGRSGSTSSASMPSPDTYNNTRYSDP | | |
| | TPLENLHGKSGIESSKNTRKQCNIGIKKGVSPSRSLANVILKNLMLYANVQHPN | | |
| | VKPDNLELVQDTLQNIQLSDNGEDNDGNSNNNDIEDGEEKDESQSYEKNNNTNL | | |
| | NRLSRHGNSLIRPSTLRASYTEEDDNEDDNDKSDASEYVNAKVEERISKIERPV | | |
| | SLNDIIEELTKISNSAGLTNDNDATITLARTISNAGSYSDKKDQPGGHVDEGIGST | | |
| gene | | | |
| CDS | | | |

SOANTLDGEFASNMPINTNTWTPERSLSRRENTYRISQOEKVEQSDVEMKND
DEERKLKNTIKVEIDPHKSPRODEENMSPPSIGFODIYHHRSSGEMMO
EMGIERKEAEVYKVRNDYVDEQDELREGTDMKPRDNDKTRKRRRNGTMMN
NKMREDNENBODDNEENVDORNELDNSKHITISLENGEKTEVSNKEENNS
TSTATSQTRKIEKTFANLRPRRHHHSDSSSPSSPSSPSPINDDAVHVRKSK
KLGKSGREPEVPIVLRNPRRHHHRSOSOKISVTKLSDSQPOOIOLOPDEBA
IEIEKRESDSESLPOLQPAVSVSTKSNSDREBEAKKKNRSMTTESINOSH
HVOKENTDEOKAOLAPAOVOYOTSVPOASAPYONASAPOTSAPEVASOTOPAP
PIKHTSLIPRKLTFADYKPKDKRNSPVOTSDSFGPPLPLTYSTYIMDHRPIV
ERATYRLSHLKSLSKRGLEQVLLSNMTAYLNLVNHNTLIMEOVADHKEQOQOQOQ
P"

BASE COUNT 1430 a 846 c 867 g 1013 t
ORIGIN

Query Match 52.9% Score 1560.4; DB 8; Length 4156;
Best Local Similarity 99.9% Pred. No. 4.5e-287;
Matches 1561; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1387 gagcctactctgaagaattttaccatgaattcctgaataatttgagtttcaggaattt 1446
|||||
Db 1 GAGCTCTACTTGAAGATTTTACCATGAAATCTCTGAATTTGGAGTTTCAGAAATTT 60
Oy 1447 tataattgcacaataacttgaggaataatttacacgtcaatgcttttgactgaactt 1506
|||||
Db 61 TATATTTGCACCAATTAAGTGAAGAAATTTTACACGTCAATGCTTTGACTAGTACTT 120
Oy 1507 aaactaataccgaacttcgcaactaagctatcaacagttatttgagcaacatcgctttc 1566
|||||
Db 121 AAACCTAATACCGCATTTGGCAACTAAGCTATCAACAGTTATTTGGCAACCATCCTTTT 180
Oy 1567 ttgttgacttgagcagcacacacatgcttatgagaatattagaagaagctctacagac 1626
|||||
Db 181 TTTTGACCTTGGCGACGACACACATGCTTATGACCAATTCAGAGGCTTCATGACAC 240
Oy 1627 tgtttccattctgctgaacaactgcttccaaatttatacaacaacacttttgagaggt 1686
|||||
Db 241 TGTTCATCTCGTGACACACATGCTTCAATTTTATACCAACACTTTTGGAGGTT 300
Oy 1687 aacgaattgtatctgtaagaagagcgaggaacactatgctgctgacataactcgtgac 1746
|||||
Db 301 AACCAAGTTTGTATTCGTAAGAACAGGCGGAGCACTATGCTGCATATATCTGCATGC 360
Oy 1747 cctttgcaataatcatgaggggtttccctgctcctcaagaattgaattacattcacagtt 1806
|||||
Db 361 CCTTTCGAATATCATGGGCTTCTCGTCTTCAAGATTGAATTTACATTCACAGTGT 420
Oy 1807 agacaagaagctggaacgcaatttcccaaatlgctcctcaactcgtgaataagtgctactgc 1866
|||||
Db 421 AGACAAAGAAAGCTGAGCAGCATTTCCAAATGTGTCAATCTGGAATATAGTGTACTTGC 480
Oy 1867 actgctgctccttggaataataatccctgaaggaataccttcaaacactcgtgtaagaactcc 1926
|||||
Db 481 ACTGCTGTCTCTTGATTAATATCCCTGAAGATPACTTACAAACTCTGTGAGAACTCC 540
Oy 1927 tggtaataagaataacccttgagccttttctgtaactgtaacggttaaatctcta 1986
|||||
Db 541 TGGTATATGAATTAACCTTTAGCCTTTTACGTAAGTGTATACCGGTTAAATTTCTTA 600
Oy 1987 tctactataaccttttctactactataatgaattcctacgagcagcgagctttgt 2046
|||||
Db 601 TGTACTATAACCTTTTCTACTACTATATGGAATTCATGAGAGGACGGGCTTTTGT 660
Oy 2047 acggaagagtgaaaaaacgagtttggttgcttggtgtaagaatttgaggaactataa 2106
|||||
Db 661 ACGGAGAGGTGAAAAAAGAGTTTGGTGTGTGGTAAAGAAATTTGGAGAGACTATATAA 720
Oy 2107 gtacctataactttgataagcagcacaataacaagtgctgctgcaatgtaagt 2166
|||||
Db 721 GTACCTAATCTTTGTATTAACGACTCAATACAACTGCTGTGCAATGATGAAGT 780
Oy 2167 tgcagatctaagagtaagaagagtgagctctaataagtttcagagcttttctttt 2226
|||||

Db 781 TGTCAGATCTAACAAGTAGAGAGAAGGTGCATCTAATAGTTTCGACGTTTCTTTT 840
Oy 2227 taagtttttatttggttctctagaatttaaggtcttaagtttggttttggtg 2286
|||||
Db 841 TAAAGTTTATTTTGGTCTCTAGAAATTAAGGCTTAAGTTAGTTTGGTTTGGTGG 900
Oy 2287 gtttacaatttcaatccaagaaggaatttagctgtcttcttataatgccaatagaat 2346
|||||
Db 901 GCTTACATATTTTCAATTCAAAGAGAAATTTAGCTGTCTTTTATATATGTCATATGAGAT 960
Oy 2347 aacgagagcagctgctgactacatacaagcgaataagcgatcgctagtcacaaggaataa 2406
|||||
Db 961 AACGAGACATGCTGCGTACTATACATCAAGCGATTAAGCGATCGTACGTAAGGATATA 1020
Oy 2407 cggaaagctgaagtttgatgctgcgaaggtcccttgcaatgaaatccgacgtaaaa 2466
|||||
Db 1021 CGGAAGTCTGAAGTTTGTATGCTGCACAGTCCCTTGCAATGAATTCGAGCGTAAAA 1080
Oy 2467 aacctaaagaattgctgattgggtlcaatggaatttactatctgaatccagaattagata 2526
|||||
Db 1081 AACCTAAAAATTTGCTGATTTGGTCAATGATTTACTTATGATCCGAATTAATATA 1140
Oy 2527 aaattcggtygggaatctagtgaggagacatgctgctgacagacatccagttctg 2586
|||||
Db 1141 AAATTCGGTGGGAATTAAGTGGAGAGATCATGCTGTGGACAGACATCCAGTTCTGCG 1200
Oy 2587 tcaatgccaagtgcacacacacacacggttaataacacagatatagatccaaactccgcta 2646
|||||
Db 1201 TCAATGCCAACTGACACACACACCGTTAATACACACATATATGATGATCAACTCCGCTA 1260
Oy 2647 gagaacttgcattggaaggggtaactcaaggaatagaatccctcaataagaactaaagaat 2706
|||||
Db 1261 GAGAACTTGCAATGGAGGGGTAATCACTCAGGATAGATATCTTCATTAAGACTTAACAAAGT 1320
Oy 2707 aactactaggtataaaaaaagtgctcactctccatccaggaatttaatgctaaagta 2766
|||||
Db 1321 AACTATAGGTATATAAAAAAGGTGTCTCATCTCATCCAGGAATTAATGCTAAGCTA 1380
Oy 2767 ttaagaataaactattatlggttctccgcaatacaacaccccttaagcttaagcctgatac 2826
|||||
Db 1381 TTAAGAAAGAACTTATATTTATGCTTCCCGCATATCAACACCTTAACGTTTAACCTGATAT 1440
Oy 2827 ttccatgagcttgtaacaagatcttatacaaatatacaaaccaagcaacatggtgaaagt 2886
|||||
Db 1441 TTCTCAAGCTTGTTCACAGATACTTTACAAATATATACAACTAACGCAATAGGTGAGAT 1500
Oy 2887 aatgatgggaatagcaatgaataaataacgatatlgaggaataatgggaggaataaataca 2946
|||||
Db 1501 AATGATGGGAATPACCAATGAATAATPACGATATGAGATATGAGGAGATTAAGAAATCA 1560
Oy 2947 ca 2948
Db 1561 CA 1562
RESULT 3
YSCNRIG LOCUS 4130 bp DNA PLN 11-DEC-1996
DEFINITION Saccharomyces cerevisiae Nrc1 gene, complete cds.
ACCESSION L42821
VERSION L42821.1 GI:857371
KEYWORDS
SOURCE
ORIGIN
Saccharomyces cerevisiae
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; Saccharomycetaceae; Saccharomyces.
REFERENCE
1. (bases 1 to 4130)
AUTHORS Bi.E. and Pringle,J.R.
TITLES ZDS1 and ZDS2, genes whose products may regulate Cdc42p in
Saccharomyces cerevisiae
JOURNAL Mol. Cell. Biol. 16 (10), 5264-5275 (1996)
MEDLINE 96413277
FEATURES
location/Qualifiers

| Source | 1. .4130 | /organism="Saccharomyces cerevisiae" |
|----------------------------|---|--------------------------------------|
| gene | /db_xref="taxon:4932" | |
| | /map="Chromosome XIIR" | |
| | 745. .3492 | |
| | /gene="Ncr1" | |
| | 745. .3492 | |
| CDS | /gene="Ncr1" | |
| | /codon_start=1 | |
| | /protein_id="AAB38423.1" | |
| | /db_xref="gi:857372" | |
| | /translation="MSNNDNESMLRTTSSDKAIAORHKKSEVLIQAOSLDFEIRSRV | |
| | KNIKRLSGSMDLIDPELDIKRGGESEGRMSCTSSSSMSPDITTYNNITXSDP | |
| | TPLENLRGRNSGDISSSKRTQGYTLCKKVSPPSRKLANVNLKNILVPAOHPP | |
| | VKPNFLELVDTQNIQLSDNGEDNGNSENNIDIDNGEDKSQSENNKENTIN | |
| | NRGRHGNASLIRSPSTLRSDYATTEFLARTLSMAGSDKSPQPOEGHVDGDIKERV | |
| | SLRITVELTKIISNSAGLTNDATITLARTLSMAGSDKSPQPOEGHVDGDIKERV | |
| | SOAULDDGEFASMPINNTWMPERSLRSRNTYRIRSDOEKVEYEDVDEKKNID | |
| | DEFLIKTKNITKYEIDPHKSPRPOODEPSSPSIGRPOIYNNHRRSSSEWED | |
| | EMGIEKAEVYVVRNDTVQDLELREGTIDYKPSATIDNKEITKRRNRNGMTWLN | |
| | NKMSREDNENOGDDEENVDQSRMELNLSKKHYISLFNGEGRTEYKNEENNSSS | |
| | TSTISQTROKIEKTFANLFRPKPHNHNDSSPSSPSSIPNDVAVRVKSSK | |
| | KLGKSGREPEVPIYLRRPRPHHHHSRHSOKISVTLKDSPOOUIPLOPLEGA | |
| | IEIKKESDESLSLPOLAPVSSSTSSNSNDRREPEAKKKRKSMTETISNOHSK | |
| | HVQENTDEOKAOLAPAOBOYOTSVPOASAAPONSAPVOTSAPIVESAOTQAPAP | |
| | PLKRTSLIPPKKTLFADYKKRDKRNSPVQFDSAFGEPLPLITVIMEDHRLIVN | |
| | ERAIYRLSHLKLNSKRGRLREOVLLSNFVAYLNLVNHITVEQVAHDKEOQOQOQ | |
| | P" | |
| BASE COUNT | 1428 a 838 c 876 g 988 t | |
| ORIGIN | | |
| Query Match | 46.2%; Score 1361; DB 8; Length 4130; | |
| Best Local Similarity | 100.0%; Pred. No. 3.8e-249; | |
| Matches 1361: Conservative | 0; Mismatches 0; Indels 0; Gaps 0; | |
| Oy | 1588 ccattcttatgagcaattacagggaaggtccatgtgcaactgtttccatctgcgtgacaac | 1647 |
| Dd | 1 ccattcttatgagcaattatcacgaagggctccatgacaaactgtttccattctgctaacac | 60 |
| Oy | 1648 atgcttccaattatatacacacacacttttggaggtttaaccgaattgtatcgttaag | 1707 |
| Dd | 61 atgcttccaattatatacacacacacttttggaggtttaaccgaattgtatcgttaag | 120 |
| Oy | 1708 aacaggcgaggaacctatggtgctgataaaccctgacgtccctttgcaatalcatgaggct | 1767 |
| Dd | 121 AACAGCGGGAACCTATGCTGCTGCATATATCCTGCATGCCCTTTCATATCATGGGCTT | 180 |
| Oy | 1768 tctgttctcttcgaagtgaattcattcatcattcaacgtatagaacaagaagctgtgacgat | 1827 |
| Dd | 181 TCCTGCTCCTTCAACATTGAATTATTCACATTAACAATAAGCAAGAAAGCTGGACGAT | 240 |
| Oy | 1828 ttccaactgttctcaactcgtgaataagtgtactctgcaactgctggtccttgatgaat | 1887 |
| Dd | 241 TTCCAAAATTTGGTCTCAATCTGGAATAAATGTGCTACTTCCGACTGCTGGTCTTGGATTAT | 300 |
| Oy | 1888 atccctgaagatccttacaaactcgtgtaggaactccctggttatagataaaccttta | 1947 |
| Dd | 301 ATCCCTGAAGATACCTTACAAACTCTGTTAGGAACTCCTGGTTATTAATAAACCTTTA | 360 |
| Oy | 1948 gccctttttagtactgtatatacgttttaaaaattcccatgtactataacttttttaac | 2007 |
| Dd | 361 GCCCTTTTACGTACTTGTATATACCTTTTAAAAATTTCCATGATATTAACCTTTTTCAC | 420 |
| Oy | 2008 tactactatggaatcctatcgagcgaccgggctttgtttagcgaagaagtgaaaaaatcga | 2067 |
| Dd | 421 TACTATTATGGAATTCATCCAGCAGCAGCGGCTTTGTGTACGGAAGATGAAAAAATGCA | 480 |
| Oy | 2068 gttctgtgcttttggtagaaagtctggaggaactaaagtaactatactttgtatagc | 2127 |
| Dd | 481 GTTTTGGTGTTTTGGTGAAGAAGATTGGAGAGCACTTAAGAATCACTTGTGATTACG | 540 |
| Oy | 2128 gactcaataacaacagtcgttcgttcagtcgtatcgtatcgaagtgttcagatctaagagtagaga | 2187 |

| DB | LOCUS | DEFINITION | ACCESSION | VERSION | KEYWORDS | ORGANISM | REFERENCE | AUTHORS | TITLE | JOURNAL |
|----|-------|---|-----------|---------|----------|----------|-----------|---------|-------|---------|
| Db | 541 | GACTCATATACAAAGTCTGTCAGTGTGATTAAGTTGTCAGATTCAGATGAGACA | 600 | | | | | | | |
| Qy | 2188 | gaagcggagatcctaaagcttcgaagcttcccttctttaaagcttctatctgctcc | 2247 | | | | | | | |
| Db | 601 | GAAAGTGGCATCTAATAGTTTCGACAGTTTCTTTTAAAGTTTATTTATTTGGCTCC | 660 | | | | | | | |
| Qy | 2248 | tgaatttaagctcttaagtttctgcttctgcttctgcttctgcttctgcttctgctt | 2307 | | | | | | | |
| Db | 661 | TGAATATTAAGTCTTAAAGTTTGGTTTGGTTTGGTTTGGTTTGGTTTGGTTTGGTTT | 720 | | | | | | | |
| Qy | 2308 | aaggagatttgaatctgcttcttcttcttcttcttcttcttcttcttcttcttctt | 2367 | | | | | | | |
| Db | 721 | ACGAGATTTTAACTGCTCTTTTAAAGTCCATTAAGATTAAGATTAAGATTAAGATTA | 780 | | | | | | | |
| Qy | 2368 | acatcaagcgaataagcgaatcgctgaatcaaaaggaataaagcgaagcgaagctt | 2427 | | | | | | | |
| Db | 781 | ACATCAGCGATTAAGCGGATTCGCTAGTCAAAAGGATTAAGCGGATTCGCTAGT | 840 | | | | | | | |
| Qy | 2428 | gctgcacagctcccttgacaaatgaatccgcgaagctaaaaaacctaagaatctgct | 2487 | | | | | | | |
| Db | 841 | GCTGCACATCTCCCTTACATGAATAATCCGACGGTAAATAAACCCTAAAGATTTG | 900 | | | | | | | |
| Qy | 2488 | gggtcaatgatttcttcttcttcttcttcttcttcttcttcttcttcttcttctt | 2547 | | | | | | | |
| Db | 901 | GGGTCAATGATTTAACTTTATTAATCAGAAATTAAGATTAATTTCCGTGGGAACT | 960 | | | | | | | |
| Qy | 2548 | gggagacgaatcagctgctgacgaacccagcttctgctgctgctgctgctgctgct | 2607 | | | | | | | |
| Db | 961 | GGGAGACGATTCATGCTGTCTGGCACAGATTCAGTCTTCTGCTCAATGCCAAGT | 1020 | | | | | | | |
| Qy | 2608 | accgcttaatacacacgaatatacgatcccaactcgcgttagaagacttgcatggag | 2667 | | | | | | | |
| Db | 1021 | ACCGTTAATTAACACAGATATATGCGATTCACACTCCGCTAGAGAACTTGATGG | 1080 | | | | | | | |
| Qy | 2668 | aactcaaggatgaatctccctccaaataagactaaacaaagtaactacttagtata | 2727 | | | | | | | |
| Db | 1081 | AACTCAGGATTAAGAACTCTCAATTAAGACTAAACCAAGTAACTACTTAAATTA | 1140 | | | | | | | |
| Qy | 2728 | gggttcaactcccaatcgaagaataatgaatgctcaagctttaaagaagaactata | 2787 | | | | | | | |
| Db | 1141 | GGGTTCACACTCCATCCAGGAAATTAATTTGCTAACGTTTAAAGAAAACTTAT | 1200 | | | | | | | |
| Qy | 2788 | gttccgcgaatcaacacccctaagcttgaagctgtaacttctctagagcttgta | 2847 | | | | | | | |
| Db | 1201 | GTTCGCCGCAATCAACACCCCTTAACGTTAAGCTTGATTAATTTCTAGAGCTT | 1260 | | | | | | | |
| Qy | 2848 | actttacaaataatacactaagcgcaaatggtggaagataatgattgggaatgca | 2907 | | | | | | | |
| Db | 1261 | ACTTTACAAAATTAATCACTTAAGCGCAATGTTGAAGATTAAGTGGAAATGCA | 1320 | | | | | | | |
| Qy | 2908 | aataacgaatctggaataatgaggagataaagaatcaca | 2948 | | | | | | | |
| Db | 1321 | AATTAACGATTTAGAGATTAATGCGAGATTAAGATTAAGATTAAGATTAAGAT | 1361 | | | | | | | |

REFERENCE 2 (bases 1 to 4225)
 AUTHORS Jia, Y., Li, J.-P. and Butow, R.A.
 TITLE Direct Submission
 JOURNAL Submitted (30-JUL-1996) Biochemistry, University of Texas
 Southwestern Medical Center, 5323 Harry Hines Blvd, Dallas, TX
 75235-9038, USA

FEATURES

source 1. .4225
 Location/Qualifiers
 /organism="Saccharomyces cerevisiae"
 /strain="COP16107"
 /db_xref="taxon:4932"
 /chromosome="XIII"
 gene 547..3294
 /gene="ZDS1"
 CDS 547..3294
 /gene="ZDS1"
 /note="gene also called RIG2S1, CES1, NRCL; Highly charged
 protein; may be a transcription regulator; multicopy
 suppressor of mrtg2-5"
 /codon_start=1
 /product="Zds1p"
 /protein_id="AAB06617.1"
 /db_xref="GI:1508861"
 /translation="MSNRDNESMLRTTSDKALASQDRKRSEVLLIAQSLDNEIRSV
 KNLRLSIGMDLIDPDLIDKEGESSGRSMGTSSTSSAMSPKDTTNNVNTYSDP
 TPLENHRGNSGIESNKTGKQNYLGIRKGVSPSKLNNANPKLNLWLPAMQHN
 VKPDNLEIYODTOLNOLSDNGEDNGNSENNIDJEDNGEDKESQYENKENTNL
 NRCLSRGNASLIRPSTLRSTIEFDNEDDDNKGSASSTYKVRERLSKIERV
 SLRDTIELTKISAGLTNDALITLRTLSMAGSYDKDQPECHYEGEDIGFST
 SOANTLDDGEASMPINNTWPERSLRSRNTYRIRKOEKVEDESDMKND
 DEERLKLTKTIKEIDPHKSPFROQDESENSPGSIGFODIYNHRKSGSEMIO
 EMGIEAEVEPVKRVNDYEDOLELREGTDMVKPSATIDNKEKTRHNRNGMTLN
 NKMSREDNENGDENENFVDSORMEIENKHYSLINFGKTEVSKKEENNS
 TSTPSTQKTIKTEPANLRPRKHHKHDSSSSPSSPSPIPNNDAYHVRKSK
 KLGKSGREPEVPLKRNPRPHRHHSRHSOKIYVTKLDSQDQIPLQPLEGA
 IEIEKKESDESSELPQAPAVSVSTSNSSNDRBEBAKKKKRSTTEISNOHKS
 HVOKENDEQKAOLOAPQEOVOTSPVQASAPVQNSAPVOTSAPEASQTOQAP
 PLKHTSILPRKLFADVKRDKPNSPVOTDSAFGPEPLITVSTVIMDHRPIVY
 ERAIYRISHLKLSKRSKRGLREQVLLSNFMAYILNLVNHITLMEQVANDKEDQOQOQOQ
 P"

BASE COUNT 1465 a 850 c 907 g 1003 t
 ORIGIN

Query Match 39.5%; Score 1163; DB 8; Length 4225;
 Best Local Similarity 100.0%; Pred. No. 1,7e-211;
 Matches 1163; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1786 gaattacattccagtagtagaacaagaagctgagcattcccaattggtcgaat 1845
 DB 1 GAATTTACATTTCCACAGTAGTAGACACAGAAAGCTGGACGCAATTTCCAAATTGGTGCAT 60
 OY 1846 ctggaataagtgctactctgcgactgctgctccttgatataatccctgaagatacct 1905
 DB 61 CTGGAATTAAGTGCTACTCTCGCACCTGCTGTCCTTGGAATTATATCCCGAAGGATACCTT 120
 OY 1906 acaaaactctgttaggaactcctgtgtatagaataaaccttaagccttttttaagctactg 1965
 DB 121 ACAAACTCTGGTAGACACTCTCGTTATAGATAACCTTATAGCCTTTTTCACGACTACTG 180
 OY 1966 catacctgttaaaattcctctgtactataaacttttcaactataatggaattcta 2025
 DB 181 TATACCGTTTAAATTTCTTCTATGTAATTAACCTTTTTCCTACTATATATGGAATTTCTA 240
 OY 2026 tcgaagcagcgggcttctgttactggaagaagtaaaaaatcgaagtttgggttcttggta 2085
 DB 241 TCGACGACGCGGGCTTTTGTGTACGGAAGAGGAAATAATCGAGTTTGGTGTGGTGA 300
 OY 2086 aagaattggaggaactataaagtaactacttctgtatataaggaactataaagaagtcgt 2145
 DB 301 AAGAAATTTGGAGGACTATAAAGTACTCTTGTGTATATGAGACTCAATTAACAAGTGTGT 360
 OY 2146 tcgtctcagtgatctgaagtgctcagatctaagaagtagagaagtgagcattcaataag 2205

DB 361 TCGGTGACGTGTAATTAAGTTGTCAGATCTTAAGAGTAGAGAGAAAGTGCATGTAATAG 420
 OY 2206 gtttcgacgtttctcttctttttaaaggttcttacttgcttccctgaattaaagtcctag 2265
 DB 421 GTTTCGACGTTTCTTTTAAAGTTTAAATTTGCTCTCTGAAATTAAGGCTTTAG 480
 OY 2266 ttaagtttggttcttctgttggtgtatataatcttaattcaaaagaagatttaagtcgtct 2325
 DB 481 TTAGTTTGGTTTGTGTTGTTGTTGTTACATATTTTCAATTCAAGAGATTTAGTGTCT 540
 OY 2326 tttaaatgttcaatagagataacgagagacatgctgcttactataatcaagcgaataagcg 2385
 DB 541 TTTATTAATGTCATATAGATTAACGAGACATGCTGCTACTACATCAAGCATTAAGCG 600
 OY 2386 atcgtctagcaagaggaataacggaagctcgaagtttcttgatctgctgcaagtccttgac 2445
 DB 601 ATCGCTACTCAAAAGGATAAAGGAAAGTCAAGTTTGATTTGCTGCACACTCCCTTCAC 660
 OY 2446 aatgaatcgcgcgctaaataaacctaaagaattgctgattggttcaatgattactt 2505
 DB 661 AATGAATCCGACGATTAATAAACCTTAATAAGATTGCTGATTGGTCAATGATTACTT 720
 OY 2506 atgatacgaattagatataaaatlcggttgaggaaatcagtgaggagacatcagtcgtct 2565
 DB 721 ATTGATCCAGATTAAGATATTAATGCTGTTGGGGAATCTAGTGGAGACATCATAGTCT 780
 OY 2566 ggcacgacatccagttcgtcgtcaatgcgaatgcacacacacccgttaataacacaga 2625
 DB 781 GGACGACATCCAGTTCTGCTCAATGCCAATGCACACACACCCGTTAATAACACACGA 840
 OY 2626 tatagcgttccaactccgctagagaacttgatagtgaggaggaactcaaggatagaatcc 2685
 DB 841 TATAGCGATCCACACTCCCTAGAGACTTGATGAGAGGGTAACTCGGATGAAATCC 900
 OY 2686 tccataaagactaaacaagtgtaactacttagtataaaaaaagtgctcaactccatcc 2745
 DB 901 TCCATATAGACTTAACAAGTAACTACTAGTATTAATAAAGGTTCACCTCCATCC 960
 OY 2746 aggaatlaaatgctaacgtatlaaagaataactatataagtggtccgcacatcaaac 2805
 DB 961 AGGAATTAATTAATGCTAAACGATTAATAAGAAACCTTATATGCTCCGCCAATCAACAC 1020
 OY 2806 cctaagcttaagcctgataattcctctagaagcttgacagaacttcaaaaatatacaa 2865
 DB 1021 CCTAACGTTAAGCCGATTAATTTCTTAGAGCTTGTAACAAGTACTTTACAAATATACAA 1080
 OY 2866 ctaagcgaacatggtgaagaataatgataagtagaatacgaataaagcgaatcgaagat 2925
 DB 1081 CTAAAGCAGATGCTGAAGATTAATGATGGAATAGCAATGAATAACGATATTGAGGAT 1140
 OY 2926 aatgggaggaataaagaatcaca 2948
 DB 1141 AATGGGAGGATTAAGAATTCACA 1163

RESULT 5
 SCU32580 3250 bp DNA PLN 24-AUG-1995
 LOCUS Saccharomyces cerevisiae Ces1p (CES1) gene, complete cds.
 DEFINITION Saccharomyces cerevisiae Ces1p (CES1) gene, complete cds.
 ACCESSION U32580.1 GI:929924
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 baker's yeast.
 Saccharomyces cerevisiae
 Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 REFERENCE 1 (bases 1 to 3250)
 AUTHORS Schwer, B., Liu, Y. and Shuman, S.
 TITLE Identification of CES1, a high-copy suppressor of ces1-ts mutations
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 3250)
 AUTHORS Shuman, S., Liu, Y. and Schwer, B.

TITLE Direct Submission
JOURNAL Submitted (27-JUL-1995) Stewart Shuman, Sloan-Kettering Institute,
1275 York Ave., New York, NY 10021, USA

FEATURES Location/Qualifiers

source

1. 3250

/organism="Saccharomyces cerevisiae"

/db_xref="taxon:4932"

gene

466. 3213

/gene="CES1"

466. 3213

/gene="CES1"

/standard_name="capping enzyme suppressor"

/codon_start=1

/product="Cesip"

/protein_id="AAJ74626.1"

/db_xref="GI:929925"

CDS

translation="MSNRDNESMLRTTSSDKAIASQDRKREVLIAQSLDNEIRSY
KRLRLSIGMDLIDPELDIKFGESGSRSGWGTSSASMPEDTTPVNTNRYSDP
TPLENLHGRCNSGIESNKTGKGYLGKKGHPSKRLNANVLTAKNLWPAOHNP
VKPDLNLEVDOTLQNLSDNGEDNGSNENNNDIDNGEKRESOSVENNTNL
NRGLSRHGNASLIRPSTLRSTYEPEDNEDDKGDSASTVYKVERISKIREY
SLRDITBEIRISAGLTDNDATITLAKLSWASTISDKDQPEGEHYDEGDIGFT
SQANTLDGEGFASNPINNTWPEKSLRSRNTYRIRSQEKEEVDKMDND
EEERLKLKNTIKVEIDPHKCPROODESEMSPSIGFODIYHMYROSSEMEQ
EMGIEKEAEVYRVNDYEDLEREGTDVYKPSATDNEKTRHRNRGWTWLN
NMSREDNEENOGDENEENVSORMELOSKHYISLENGEKTEYSNKEENNS
TSTASOTROKIEKTPANLFRPKRHHKHDASSPSPPSPSTPNDAYHVRKSK
KLGKSGREPEPIVLRNRRPRHHRHSHSOXISVTKLDSPOOPILOPDEGA
IEIKKESDESJLQLOPAVSSTSKNSKDRKEEEKKNNKRSNTSEISNOQHK
HVQENDEDEKAOLOPAQEOVOTSVPAQNSAPVOTSAVSAVSAOTQAPAP
PLKHTSLIPRKLTFADVKPKDKPNSPVOFDSAFGLPLITVSTVIMEDHRLPIV
ERAIYRSLHLKLSNKRGLREQVLSNMFAYLNLVNHITLYMEQVADHDKEQOQOQOQO
P"

BASE COUNT 1164 a 658 c 690 g 738 t
ORIGIN

Query Match 36.3%; Score 1069.4; DB 8; Length 3250;
Best Local Similarity 99.8%; Pred. No. 1.1e-193;
Matches 1081; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

Oy 1866 cactgcgtgctctgagtaataatccctgaaggaataccttaacaacctgtaggaatc 1925
Db 1 CACTGCTGCTCTTGG-TTAAATATCCCTGAAGGATACCTTACAACCTCGTAGAAGATC 59
Oy 1926 ctggtatagaataaccctttagcctttttagctactgtataccgtttaaatcttcc 1985
Db 60 CTGGTTATAGATAACCCCTTAGCCCTTTTACGTACTGTATACCGTTTAAATTTCT 119
Oy 1986 atgtactaaaccttttcaactactatataatgaatctatcgagcgagcggtcttgc 2045
Db 120 ATGTACTATAACCTTTTCACTACTATTATGAGATTCTATCGACGACGCGGCTTTG 179
Oy 2046 taaggagaagtgaaaaaactcgagtttgcgttcttggtgaaagaatttggaagactata 2105
Db 180 TACGGAAGAGTGAAAAATCGAGTTTGTGCTTTGTGTAAGAAATTTGAGAGACTATA 239
Oy 2106 agtacctacttgcttaacgagactcaatacaagtcgttcgttcggtgtagtgag 2165
Db 240 AGTACTTACTTGTATATAGGACTCAATACAGTGTCTGCTGCTAGTGTATGAG 299
Oy 2166 ttgtcagaatctaaagtagaagaagtgagcatcaataagtggttcgagctttcttct 2225
Db 300 TTGTCAAGATCTAAGAGTAGAGAGAGTGCGCATCAATAGGTTTGAGCTTTCTTTT 359
Oy 2226 ttaaggttcttacttgctcctcctagaatttaaggtcctagtagtttggttgc 2285
Db 360 TTAAGGTTTATTTGCTCTCTAGAAATTTAAGGCTTAGTGTGTTGGTTGTTGTT 419
Oy 2286 gggttaacatatttcaatcaagaagaatttagctgctcttataaagtccaatagaga 2345
Db 420 GGGTTACATATTTCAATTCAAGAGAGAAATTTAGCTGCTTTTATTAATGCCAATAGAGA 479

Oy 2346 taacgagagacatgctgctgactacaacgagataaggcgaatcgctagctaaaggatata 2405
Db 480 TTAACGAGAGCATGCTGCTGCTACTACTACATCAAGCGATTAAGCGATCGTACTCAAGGATTA 539
Oy 2406 acggaagctcgaagtttgatgctgcacagtcctcctacaaatgaatcccgagatata 2465
Db 540 ACGGAAGCTGATGATTTGATTTGATTCGACACAGTCCCTTACATGAATAATCCGACAGGTAA 599
Oy 2466 aaacctaaagaattgctgattgggttcaatgagattactatlgatccagaattagatat 2525
Db 600 AAACCTAAAGATGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 659
Oy 2526 aaattcgttgaggaatcgaatcgaatcgaatcgaatcgaatcgaatcgaatcgaatcga 2585
Db 660 AAATTTGCTGGGATCTAGTGAGGAGATGATGATGATGATGATGATGATGATGATGATG 719
Oy 2586 gtcaatgccaagtgcac 2645
Db 720 GTCAATGCGCAAGTGAC 779
Oy 2646 agagaacttcgaatgagaggaatgaactgaagataagaaatcccaataagactaaagaag 2705
Db 780 AGAGAATCTGATGAGGAGGAGGATCAAGGATGATGATGATGATGATGATGATGATGATG 839
Oy 2706 taactactgaatgaataaaagaagtgctcacttcctcactcgaagaaatgaatgaatga 2765
Db 840 TTACTACTTATGATTAATAAAGGCTTCACTTCTCATCCAGANATTAATGCTAACGT 899
Oy 2766 attaaagaaaaactatataatggttcgcccaatcaaaccttaacgttaacgttaata 2825
Db 900 ATTAAGAAAACTATATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 959
Oy 2826 ttctctgaagctgtgacaaagatccttacaataataaactaagcgaatggtgaaga 2885
Db 960 TTCTCTGAGCTGTGACAGATCTTACAAATATACAACTAAGCGAATGATGAGGA 1019
Oy 2886 taatgatggaatagcaatgaataaataacatataatgaatgaatgaggaatgaagaatc 2945
Db 1020 TAATGATGGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1079
Oy 2946 aca 2948
Db 1080 ACA 1082
RESULT 6
SCU63849
LOCUS SCU63849 3525 bp DNA PLN 11-MAR-1997
DEFINITION Saccharomyces cerevisiae zdsip (ZDSI) gene, complete cds.
ACCESSION U63849
VERSION U63849.1 GI:1498489
KEYWORDS
SOURCE
ORGANISM baker's yeast.
Saccharomyces cerevisiae
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetaceae; Saccharomycetaceae; Saccharomycetes.
REFERENCE 1 (bases 1 to 3525)
AUTHORS Ma,X.J., Lu,Q. and Grunstein,M.
A search for proteins that interact genetically with histone H3 and
H4 amino terminal uncovers novel regulators of the Swi1 kinase in
Saccharomyces cerevisiae
Genes Dev. 10 (11), 1327-1340 (1996)
JOURNAL MEDLINE 96249381
REFERENCE 2 (bases 1 to 3525)
AUTHORS Ma,X.J. and Grunstein,M.
Direct Submission
TITLE Submitted (15-JUL-1996) Biological Chemistry, Univ. of California
JOURNAL at Los Angeles, 405 Hilgard Ave., Los Angeles, CA 90095, USA
FEATURES Location/Qualifiers
source 1. 3525
/organism="Saccharomyces cerevisiae"
gene /db_xref="taxon:4932"
315. 3062

CDS

/gene="ZDS1"
315..3062
/function="suppressor of HSL1 and HSL2; transcription
regulator"
/note="ORF YW8156.15c; CEST, see GenBank Accession Numbers
Z49260 and U32580; similar to Mslp encoded by GenBank
Accession Number U32938"
/codon_start=1
/product="Zdsip"
/protein_id="AAB49281.1"
/db_xref="GI:1498490"
/translation="MSNRDNEMLRTTSDKAIASQDRKRSEVLIQAQSLDNEIRSV
KNEKLSIGSMILLIDELIDKEGEGSRWSGTTSSASMSPTDITVNNRYSPP
TPLENHGRNSGIESNKTROGVLGKRGVSNKTAANLYKLLVWVNOHNP
VKPDNPLEIYODPLONTOLSDNGEDNNGNENNDIDNEKESQSEYKENTITL
NRLSRHGNASLIRPSTLRSTTERPDNDKDDKQDQPECHYDEGIDGST
SLADITEELPKINSAGLTNDIAITLARLISAGSYSDKDDQPECHYDEGIDGST
SOANTLDGEGFASNPINNTMTWPERSLRSPFNTRYRISOEKEKEVEDKMD
DEERLKLTKNTIKYIDPHKSPROODEENMSPSIDGFDIYHRYSGEMEO
EMGIERKEEVPYKVRMDYEDOLELREGTDMVKPSATDNEKTRHRNRGNTWLN
NMSREDNEENGGDDNENNVDSORHELDNSKKHITSLNNGEKTEVSKERNNNS
TSTATSQTRQKTEKTPANLFRKPRHHRHDSVSSPSSPSPSPINDAVHVRKNS
KLGNSGSEVEPIVLNRPRPHHRHRSQKISVYTKLSDQOQIDLPQLEGA
IEIEKESDESSESLPOLQPAVSYSSTNSRDRDEEBAKKRNRSTTEISNOHSA
HVOKEKTEDEKAOLOAPVOVOTSVYONAPONAPOTSAVAPASATQAPAP
PLKHTSLPRLKTFADYKRPDKPNSPVOTDSAFGEPLITVSTYIMEDHRLPIV
ERAIYRSHKLSNRKRLREYVLSNFMATYLVNHTLMEDVADHKEDQ000000
P"

BASE COUNT 1254 a 711 c 760 g 800 t
ORIGIN

Query Match 31.6%; Score 931; DB 8; Length 3525;
Best Local Similarity 100.0%; Pred. No. 2,3e-167;
Matches 931; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2018 gaattcctacgagcagccggcctttgttaccgagagtgagaaatcgaagttttgtgt 2077
|||||
Db 1 GAATTCATCGACGACGGCGCTTTGTTACGGAAGATGAAANAATCGAATTTGGGT 60
OY 2078 ttgtgtgaagaatttggagagactataaagactatattctgttactacgagaccataa 2137
|||||
Db 61 TTTGGTGAAGAATTGGAGAGACTATAAAGTACCTATTGTGATTCGAGACCAATAA 120
OY 2138 caagtcgtcgtcgtcagtgatgaaagtcgacatcgaagtagagagaagtgagca 2197
|||||
Db 121 CAAGTCGTTCTGTCAGGTGATGAAAGTGTGAGATCTAAGAGAGAGAAGGTGCA 180
OY 2198 tctaagagttcgaagctttctcttttaagtttttattatgtgctccgaataataa 2257
|||||
Db 181 TCTAATAGGTTTCGACGTTTCTTTTAAAGTTTATTTAGTCTCCAGAAATTTAA 240
OY 2258 ggtctagtagtttggtttctgtgtgtacatatttcaattcaagaagaaatt 2317
|||||
Db 241 GGTCTAGTAGTTTGGTTTCTTTTGGGTTACATATTTCATTCAAAGGAATTT 300
OY 2318 agcgtctcttataatgctcaatagagataagagagacatgctgctactacatcaagc 2377
|||||
Db 301 AGCGTCTTTTATATGTCACATAGAGATAACGAGAGATCTCGTCAATCAAGCGC 360
OY 2378 ataagggagcagctcagtcgaagagataagagagagctcgaagtttgaattgctcagcgt 2437
|||||
Db 361 ATAAGGCGATCGCTAGTCAAAAGGATTAACGGAAGTCTGAAAGTTTGTATTCCTGACAGT 420
OY 2438 cccttgacaatgaaatccgcagcgtlaaaaaacttaaaaaatgctcgaattggtcgaatg 2497
|||||
Db 421 CCCTTGACAAATGAATCCGACGCTAAAAAACCTAAAAAGATGTCGATTGGGTCAATGG 480
OY 2498 attactatgacccagaaatagataataaaatccggtggggaatcagtgaggagaagat 2557
|||||
Db 481 ATTACTATTGATTCACAAATTAATATAAATTCGCGTGGGGAATCTAGTGGAGAGACAT 540

OY 2558 catgtctgacgacacatccagttctgctgctacatgccaagtgcacaacacacgttaata 2617
|||||
Db 541 CATGCTCGGACGACACATCCAGTTCTGTGGTCAATGCCAAGTAGACACAACCCGTTATA 600
OY 2618 aacacagatagacgacatcccaactccgctagagaacttgcattgaggggttaaccagga 2677
|||||
Db 601 ACACACGATATAGCATCCAACTCCGTAGGAACCTTGCAAGGAGGATTAACACAGGA 660
OY 2678 tagaatctccatagaagactaaacaagaattactcttaggtatataaaaaaagtgttact 2737
|||||
Db 661 TAGAATCTCCAAATAGACTTAACAGGATTAATCTTAAGGTATTAATAAAGGTTCACCT 720
OY 2738 ctccacacaggaatataatgctaacgtattataagaataaacttataatggtccgca 2797
|||||
Db 721 CTCATCCAGGAATTTAATCTTAACGATTAAGAAATACTTTATGCGTCCGCCA 780
OY 2798 atcaaacaccttaacgttgaagcctgataaattctcctagagcctgtacaagaacttaca 2857
|||||
Db 781 ATCAACACCCCTTAACGCTTAAGCTGATTAATTTCTAGAGCTTGTACAAAGATCTTACAAA 840
OY 2858 atatacactaagcagacatgctgataagataatgataagcagacagaaataacata 2917
|||||
Db 841 ATATCAACTTAAGCAGCAATGCTGAAGATTAATGATGGAATAGCAATAAATACGATA 900
OY 2918 ttgagataatggggagagataaagaatcaca 2948
|||||
Db 901 TTGAGGATTAATGGGAGAGATTAACATACACA 931

RESULT 7
YSCBUL1
LOCUS YSCBUL1 3240 bp DNA 10-FEB-1999
DEFINITION Yeast BUL1 gene for ORF, complete cds.
ACCESSION D50083
VERSION D50083.1 GI:773189
KEYWORDS BUL1.
SOURCE Saccharomyces cerevisiae DNA.
ORGANISM Saccharomyces cerevisiae
REFERENCE 1 (bases 1 to 3240)
AUTHORS Kikuchi, Y.
TITLE Direct Submission
JOURNAL Submitted (07-APR-1995) to the DDBJ/EMBL/GenBank databases. Yoshiko
Kikuchi, Graduate School of Science, The University of Tokyo,
Department of Plant Sciences, 7-3-1 Honjo, Bunkyo-ku, Tokyo 113,
Japan (Tel:03-3812-2111(ex.4466), Fax:03-5684-9420)
REFERENCE 2 (bases 1 to 3240)
AUTHORS Yashiroda, H.
TITLE BUL1, a new component of the ubiquitination pathway in
Saccharomyces cerevisiae
JOURNAL Unpublished (1995)
REFERENCE 3 (sites)
AUTHORS Yashiroda, H., Oguchi, T., Yasuda, Y., Toh-E, A. and Kikuchi, Y.
TITLE BUL1, a new protein that binds to the Rsp5 ubiquitin ligase in
Saccharomyces cerevisiae
JOURNAL Mol. Cell. Biol. 16 (7), 3255-3263 (1996)
MEDLINE 96251273
FEATURES
source
Location/Qualifiers
1..3240
/organism="Saccharomyces cerevisiae"
/db_xref="taxon:4932"
/chromosome="13r"
181..3111
/gene="BUL1"
/codon_start=1
/protein_id="BA08787.1"
/db_xref="GI:773189"
/translation="MAKDLNDGSPPKRNPRLRPPRSDFTSNTSMANANTGRGR
OKOEGKSSRPSLSHSPKMSASATGILGIRBHLASHASTGTCPAGNSP
LRSTANATFETGRSLTDDIDINNVAVLPSPFMYNLRIHRIPOGVNDDPDRHDFPPSY

OEANNSTATGAGSSADLSHQSISTDLAGATRRSSSTNSLENLPLRTEHHSIAHOST
 AVEEDSLDIPILDDNDNFIDKLYTLPRKMTPEIITIKTKHAPIPHVAREEBS
 ILKETEYSGDLHGFIITENKSOANLKFEEMVYLSIIDKXSKRTIKRPLQIOLD
 LSASMSYSKIALSGVDFIPADVDYDSVGLNNSRVLPEGVKKKFFIKRPLQIOLD
 VTCKOEHFSCILPSPFGIDKIRNNCKYSIGIKNVRVIGCGHIGKSPILINDMSDN
 LSIINYTDARIIVGCDOKASKLYIKEREYNNRVIPFEPDANVYGERTMSQLNDITKL
 VOERLDLARKIFORLEKKEPTINRDIGHADLSGTIDDSIESDQELQRLDOLHIKN
 RNNYLVNNDLKGHDLDNGRSGNSGHNDSRAMGPFVSEELKYLKKNSSNSFLN
 FSHFLNSSSSSSNAGKNNDLGNKERTGLIIVKAKIPROGLPYAPSLIRKTN
 VFESKSHDQENNVRLSELIPEDVKRPLEKLDLOTIESDNLPHRPEIOSITTEL
 ICTTAKSDNSIPIKINSELMMNKREKLSIALDODRISKICEETKFNKPLEINLEY
 NMNGDRPRKELKFTDFITSOLENDIESICNLKVSNNLSNIFKKOYSTLKQSKHAL
 SEDSISHTGNGSSSPSSASLTPTVSSSKSLFPLSGSSSTSLKFTDOIYHKWRIAP
 LOYKRDINVLNLEFNKDIKETLIPSFESCLOCRRFYCVAMIKFENHLGVAKIDIPISVR
 OYTK"

BASE COUNT 1047 a 648 c 643 g 902 t
 ORIGIN

Query Match 26.7%: Score 786.8; DB 8; Length 3240;
 Best Local Similarity 99.7%: Pred. No. 6,1e-140;
 Matches 788; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 tgaactgttgaaacaagaagctgacaagatcaaaagcttgaagatgattcca 60
 Db 2451 TGAAGCTGTGATGAAACAAGAGAGAGTCAAGCATCAAGCTTGTACGATGATTTCCA 2510
 OY 61 ttcaaaaatttgtaaatgaaacaaagttcaaaagatttcttgatataatgaatt 120
 Db 2511 TTCAAAAATTTGTGAATATGAAACCAAGTTCACCAAGAAATTTCTTGAAATTAAGAGATT 2570
 OY 121 atataataatgaatgaagagacgctgagcacaagaagaaattacagattttatcac 180
 Db 2571 ATATATATGTGAATGAGGAGACCGTAGGCCAAGAACTGAATTTTACAGATTTTATTTAC 2630
 OY 181 ttcaagctgttgaaagatcgaagaagcattgcaacttgaagttagttgccaactt 240
 Db 2631 TTCAAGCTGTGTAACGATATGAAAGCATTTGCAACTTGAAGATTAGTTGTTACAAACTT 2690
 OY 241 atccaacattttaaataaagagctgacagcaaaccaaacacacccaagcagcattac 300
 Db 2691 ATCCACACTTTTAAATAAACAGGTGACCTTAACCAACACATCAAGACGCACTTATATTC 2750
 OY 301 tgaagatcacaatcgcacacaggttaacgtagttcaccgctgcacgctac 360
 Db 2751 TGAGGATTCATATCGCACACAGGTAAAGGTAGTTTCATCGCCCACTTCAGCGTCAATT 2810
 OY 361 aagcagtaactcttcaaccaagtagttacatttaccctagcggtagctgctac 420
 Db 2811 AACGCACTTAACCTTCATCCAAAGAGTATTTATTTTACCTAGCGGTACCTGCTTAC 2870
 OY 421 ttccctgaatttcaagcagcagattgttcataaataagttgattgcttcttcaagta 480
 Db 2871 TTCCCTGAAATTTACAGCCAGATTGTCATTAAGGTTAGGATTGCTCTTTACAGATA 2930
 OY 481 caaacagacataatgtaagacttgaaatcaataaagacataaagaaacttcaatcc 540
 Db 2931 CAACAGACATTAATGGAAGCTTGGAATTAATTAAGACATTAAGAACTTTAATTC 2990
 OY 541 aagtttgaaagctgctcagttgtaagtttattctgctgcagtaagatlaaattga 600
 Db 2991 AAGTTTGAAGCTGCTATGTTAGGTTTATTTGCTTGAGTAATGATTAAATTGGA 3050
 OY 601 aaacacattctggctagcgaagattgataccctattctgctgagcagtgacaata 660
 Db 3051 AAACCATCTTGGCGTAGGAGATTGATATCCATTCTTTAGGCAAGTACAAATA 3110
 OY 661 aaaaaacattgaaaaaaattctcgttactttcttatagataagataagattggtt 720
 Db 3111 AAAAAACATTGAAAAAATTTCTGTTACTTTCTTATAGATATAGATATATGATGTC 3170
 OY 721 gctatagatgaaggtattatcgcgtctcttgatattccattataataaattctt 780
 Db 721 gctatagatgaaggtattatcgcgtctcttgatattccattataataaattctt 780

Db 3171 GCTATATAGATGAGGATATTATTCGCGTCCTTGCATTCCTATATTATAAATATCTTT 3230
 OY 781 taatatgat 790
 Db 3231 TAAATGAT 3240

RESULT 8
 LOCUS SCRD1GEN 3020 bp DNA 04-JUL-1995
 DEFINITION S.cerevisiae rds1 gene.
 ACCESSION X88901
 VERSION X88901.1 GI:895906
 KEYWORDS rds1 gene; respiration deficiency suppressor.
 SOURCE Baker's yeast.
 ORGANISM Saccharomyces cerevisiae
 Saccharomyces cerevisiae
 Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 Saccharomycetales; Saccharomycetaceae; Saccharomyces.

REFERENCE 1 (bases 1 to 3020)

AUTHORS Stein,T.
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 3020)

AUTHORS Stein,T.
 JOURNAL Direct Submission
 TITLE Submitted (21-JUN-1995) T. Stein, Universitaet Dueseldorf,
 Botanisches Institut, Universitaetstr.1, D-40225 Dueseldorf, FRG

COMMENT Related sequence D50083.
 FEATURES
 source
 Location/Ouallifiers
 1..3020
 /organism="Saccharomyces cerevisiae"
 /strain="GRF8"
 /xref="taxon:4932"
 /db_xref="taxon:4932"
 /clone_lib="YCP50 genomic genebank"
 /clone="pUC19/YEp352"
 32..2962
 /gene="rds1"
 /gene="rds1"
 /codon_start=1
 /product="respiration deficiency suppressor"
 /protein_id="CA61363.1"
 /db_xref="GI:895907"

gene
 CDS
 /translation="MAKDLNDGPPPRKRLRPORDFPANSSTMMNVNANTRGGR
 OKOEGKSSRSPLSPKSWISSASATGILGRLPLASHHABSTGPPAGNSP
 LRSTANATPEVGRSLTDDINNVDVLPSEFNTLHRIHIOGVADPRDPPSY
 OEANNSTATGAGSSADLSHQSISTDLAGATRRSSSTNSLENLPLRTEHHSIAHOST
 AVEEDSLDIPILDDNDNFIDKLYTLPRKMTPEIITIKTKHAPIPHVAREEBS
 ILKETEYSGDLHGFIITENKSOANLKFEEMVYLSIIDKXSKRTIKRPLQIOLD
 LSASMSYSKIALSGVDFIPADVDYDSVGLNNSRVLPEGVKKKFFIKRPLQIOLD
 VTCKOEHFSCILPSPFGIDKIRNNCKYSIGIKNVRVIGCGHIGKSPILINDMSDN
 LSIINYTDARIIVGCDOKASKLYIKEREYNNRVIPFEPDANVYGERTMSQLNDITKL
 VOERLDLARKIFORLEKKEPTINRDIGHADLSGTIDDSIESDQELQRLDOLHIKN
 RNNYLVNNDLKGHDLDNGRSGNSGHNDSRAMGPFVSEELKYLKKNSSNSFLN
 FSHFLNSSSSSSNAGKNNDLGNKERTGLIIVKAKIPROGLPYAPSLIRKTN
 VFESKSHDQENNVRLSELIPEDVKRPLEKLDLOTIESDNLPHRPEIOSITTEL
 ICTTAKSDNSIPIKINSELMMNKREKLSIALDODRISKICEETKFNKPLEINLEY
 NMNGDRPRKELKFTDFITSOLENDIESICNLKVSNNLSNIFKKOYSTLKQSKHAL
 SEDSISHTGNGSSSPSSASLTPTVSSSKSLFPLSGSSSTSLKFTDOIYHKWRIAP
 LOYKRDINVLNLEFNKDIKETLIPSFESCLOCRRFYCVAMIKFENHLGVAKIDIPISVR
 OYTK"

BASE COUNT 982 a 611 c 598 g 829 t
 ORIGIN

Query Match 23.9%: Score 703.8; DB 8; Length 3020;
 Best Local Similarity 99.6%: Pred. No. 3,7e-124;
 Matches 716; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

OY 1 tgaactgttgaaacaagaagctgacaagatcaaaagcttgaagatgattcca 60
 Db 2302 TGAAGCTGTGATGAAACAAGAGAGAGTCAAGCATCAAGCTTGTACGATGATTTCCA 2361

Qy 61 ttcaaaaatttgtagaata tgaacaacaaagttcaacaagaattttcttgaaatlaaatgagt 120
|||||
Db 2362 TTCAAAAATTGTGTAATATGAACAAGTTCACAGAAATTTCTTGAAATTAATAGAGTT 2421
Qy 121 atataatgaatgaagggaagaccgttaggcacaaaggaactgaattacagatttactaac 180
|||||
Db 2422 ATATATATATGAATAGGAGAGACCGTAGGSCAAAGGAACTGAATTTACAGATTTATTTAC 2481
Qy 181 ttcaagctgttttaacgagatccgaagacatttgcaacttgaaagttaagtgttcaacaatt 240
|||||
Db 2482 TTCACAGCTGTTTAACGATATCGAAGAAAGCATTTGCAACTTGAAGTTAGTGTTCACAACTT 2541
Qy 241 atccacacatttttaaaaaaacaggtcagtaacccttaaacacacactaaagcaagcactaac 300
|||||
Db 2542 ATCCAAACATTTTAAAAAACAGGTGAGTACCTTAACACACACTTAAGGACAGCATTTATC 2601
Qy 301 tgaagatcaatataatgcacacaggaagtaacgtagttcatcgtgcagcttcaagcagtcact 360
|||||
Db 2602 TGAGGATTCATATATCGCACACAGGTACGAGTTCATATCGCCAGTTCAGCGTCACTT 2661
Qy 361 aagccagatgaactttctatccaaagagtagttatttttactaagcagtagctgtctac 420
|||||
Db 2662 AACGCCAGTAACTTCTTATCCAAAGATGATTTATTTTACCTACCGGTAGCTCGTCTAC 2721
Qy 421 ttccctgaatttacaagaccagatgttcataaatgagttagagattgtctcttcaagta 480
|||||
Db 2722 TTCCCTGAATTTTACAGACAGATGTTCATTAATAGGTTAGAGATTGCTCTTTACAGTA 2781
Qy 481 caaacgagacatlaattgtgaacttgaaatttaataagaacattaaagaaacttaattcc 540
|||||
Db 2782 CAACGAGACATTAATATGCAACTTGGAAATTAATTAAGACATTAAGCAAACTTTAATTC 2841
Qy 541 aagttttaaaggtccctcaatgttgaagttttatgtcgttcgtagaagaatgaattga 600
|||||
Db 2842 AAGTTTAAAGCTGCCATATGTGATGATTTATGCTTCAGATTAAGTAAATTTGA 2901
Qy 601 aaacatcttgagcgaagaaatgatataccctattctgttgaagcaagtacaaat- 659
|||||
Db 2902 AAACCATCTTGCGCGTACGGAAGATGATATCCCTATTTCTGTAGCAAGTACAAATA 2961
Qy 660 aaaaaaacattagaanaaaattctcgttactttcttataagatatagatatgtatgct 718
|||||
Db 2962 AAAAAAACATTAAGAAAAATTTCTCTTACTTTTCTTATAGATATATATATATATGCT 3020

RESULT 9
YSCGENC58X 270 bp DNA 11-MAR-1996
LOCUS YSCGENC58X 270 bp DNA 11-MAR-1996
DEFINITION Yeast Eco RI fragment.
ACCESSION M87379
VERSION M87379.1 GI:1220615
KEYWORDS
SOURCE Saccharomyces cerevisiae DNA.
ORGANISM Saccharomyces cerevisiae
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; Saccharomycetaceae; Saccharomyces.
REFERENCE 1 (bases 1 to 270)
AUTHORS Davies, C.J., Trgovcich, J. and Hutchison, C.A. III.
TITLE Limited application of a random DNA sequencing strategy to the
Saccharomyces cerevisiae genome identifies 3 new genes and
indicates that approximately 90% of the genome may be transcribed
JOURNAL Unpublished (1994)
FEATURES
source location/Qualifiers
1..270
/organism="Saccharomyces cerevisiae"
/db_xref="taxon:4932"

BASE COUNT 67 a 30 c 71 g 102 t
ORIGIN
Query Match 7.6%; Score 225.2; DB 8; Length 270;
Best Local Similarity 94.1%; Pred. No. 4e-33;
Matches 256; Conservative 0; Mismatches 13; Indels 3; Gaps 2;

Qy 2018 gaattctacgagcg-accggagctttgttaccgaaagatgaanaaatcagagtttggctg 2076
|||||
Db 1 GAATTCATTCGAGACACCGGAGATTTGTATACGAGAGCTGAAAA--TCGCGTTGGTG 58
Qy 2077 ttgtgtgaagaatttgaagacataaagtaacctatacttctgtattacagactcaata 2136
|||||
Db 59 TTTGTGTAAGAATTTGAGAGCATATAAGTACTATACCTTTATTTACGGGGTCAATA 118
Qy 2137 acaagctgtcgtgtcagatgtgatataagttgtcagatcctaaagtagaagaagtgagc 2196
|||||
Db 119 ACAAGTCTGCTGTCAGATGTGATGAAGTGTCTAGATCTAAGAGTACAGAAAGCTGCG 178
Qy 2197 atctaaatgatttcagagctttcttctttaaaggcttttattgtctccagaactta 2256
|||||
Db 179 ATCTAATAGTTTCACAGCTTTTCTTTTATTAAGGTTTATTTATTTGGTCTCTAGAAATTA 238
Qy 2257 aggtcttagttagtttggcttggcttggc 2288
|||||
Db 239 AGGTCTTAGTTAGTTTGGTTTGGCTTTGGTGG 270

RESULT 10
SC8339 27559 bp DNA 11-AUG-1997
LOCUS SC8339
DEFINITION S. cerevisiae chromosome XIII cosmid 8339.
ACCESSION Z49210 Z71257
VERSION Z49210.1 GI:798881
KEYWORDS aminoglycoside resistance; APR1; DAT1; MDH1;
oligo(dA)/oligo(dT)-binding; orotate phosphoribosyltransferase;
SEC65; signal recognition particle; SNO1; spore germination; URA5;
VAN1; vanadate resistance.
SOURCE baker's yeast.
ORGANISM Saccharomyces cerevisiae
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; Saccharomycetaceae; Saccharomyces.
REFERENCE 1 (bases 1 to 27559)
AUTHORS Skelton, J. and Churcher, C.M.
TITLE Unpublished
2 (bases 1 to 27559)
AUTHORS Barrell, B., Rajandream, M.A. and Walsh, S.V.
REFERENCE Direct Submission
TITLE Submitted (02-MAY-1995) Saccharomyces cerevisiae chromosome XIII
sequencing project, Sanger Centre, Hinxton Hall, Hinxton, Cambridge
CB10 1RQ E-mail: barrell@sanger.ac.uk
COMMENT
Notes:
All CDS over 100 codons have been analysed. CDS that are completely
overlapped and those that are overlapped by more than 50%
of their length by a larger CDS have been omitted from this
analysis.
Details of the omitted CDS are available on request. The more
significant matches with motifs in the PROSITE database are
also included but some of these may be fortuitous. The length in
codons and the calculated codon adaptation index (CAI)
is given for each CDS.
Cosmid 8339 is overlapped at the start by lambda 7056 and at the
end by cosmid 9514, EMBL X80835.
FEATURES
source location/Qualifiers
1..27559
/organism="Saccharomyces cerevisiae"
/strain="AB972"
/db_xref="taxon:4932"
/chromosome="XIII"
/clone="cosmid 8339"
/map="13L"
/note="YM8339.01, unknown, len: 366, CAI: 0.11, overlaps
and extends YM7056.08"
/codon_start=3
/product="unknown"
/protein_id="CA89100.1"
/db_xref="GI:798882"
/db_xref="SWISS-PROT:Q03210"

misc_feature

1.76 "overlap with lambda 7056"

CD5

/note="YM8339.02, unknown, len:1134, CAI: 0.14; P500017
1495..4899
/note="ATP/GTP-binding site motif A"
/codon_start=1
/product="unknown"
/protein_id="CA89101.1"
/db_xref="GI:798883"
/translation="MSNSKKKPVANVYVYRQOQDYNGMNAVGNPMVHPADVFNGAG
OYGPQHPAYVNSPLPNIPTPTPTAGALFPSHLMGSPVPSNMQSGYASAS
SLUKKATSRPVSNNNGNGNSNQNTNGAVTPSYVTRMGKNSRSNNNTNRYAT
NNHNGGQTRNNSGRRTEFARNNTEDDILPEMLDORPCINKVLPTDDAARTSLI
EWHNSDILSHIVKNFVKSNTLESAYLLEGKSDSKDVERKNLSLISFLTKGDCLN
FNNALIRSEKRTFLKSEALNFKVCVANTPKCLPTFESEALTENAEEDIDNGST
MNSALHHNIKAKDTRSLITIEFKSPKSDLPKLFQDLSRKNKRYLIESIDLVNT
DYPNSQEPENAVAYLPELINSMAIEYDLKRYKSLGSKPEYSLAPLYVSSRSVV
ANIYKGTSTHRLSVSVTAGNNDSNNNGNNSNSGTLTNNNSISGVSYGHSN
MLSTLSVSSVSLNEIDMLATKQCEVLDGYLEINRDVOTPIEBHSHLSVWKLS
KTTENSROFQDIPSPALPLENHEFMNDSQNGAIIIPOLLATPSPVPSNMQRPVLA
PPIPTQESQENFNVSAKVAASGSDIGRTYIGINIPRSKAEDICNVVCGILDSIT
LYPEKKIEVFTELEAPSAVOFYANGSFIDPVLKGNLKVCGKSGPLPKLSIATVNT
GASRVVYSLPEPAFKKEFHIDPOKTKHETLSLPDAEQLEDETSYGDIDQINLYSD
SHCCINIFNNISAIISLVEEMKEESTVONESGYTLKRAIEKFGGRKGLLINYGD
RCGNLNNKLIAKNGRFRYKVRPSYNTLRSLKEKRQNDIDDKAEFKPLMESEL
GSLDAHNDGGGETAGNTANGHEMESLEENENGEKNTGSGGLAVALSSDYKRAAT
SBETVEDYFNKSSGSDSDSDEYVIMSPDEPAALKSOTLRSSQOTVNSKRPVKI
DEEEVAYQMSDILNYSRSLROAPRAPPSLTLSNHSKNNETPMQDIFTEGFTANNKKRR
GSEFAHRTIRPGSDVAOYLAQVQSTFMYAANILGASAEQNTHPDE"
2845..2868
/note="P500017 ATP/GTP-binding site motif A"
5448..7076
/gene="ATRI"
5448..7076
/gene="ATRI"
/note="YM8339.03, ATRI gene (SNQ1), len: 542, CAI: 0.11,
ARI1_YEAST P13090 aminotiazole resistance protein"
conflict at C-terminus, integral membrane protein"
/codon_start=1
/product="Atr1p"
/protein_id="CA89102.1"
/db_xref="GI:798884"
/translation="MGNSQVLAETSEKVEYENETELPYKSSRDNNIGESLTAFAFQ
SDHEWVSNQKQNNPNFYKVAOEYLFETQMISOLLNAQOTPTLISNLTSSPSS
EENSQMSLMSPELVSGSESLISGRGIDYELKMLLVGYLYVIMSLIGITRYS
DTEFLTSARQGLGIAFLPNVLGTLGITYIGVFRKRTVYSGAPAAIGATIGCLFL
AGLIGTEDPKQMPAEVAVYSIAAFINFLVSYALPSTPINIHFSMDWIGSVYGVIG
LILLNENQAPISGMOQAYIIVILISIVLAVFYIYEIRFAKTPLLPRAVIDRHM
IOIMALEFGMGSGFGLFTFYEFQOLNIRQYALMAGGYEFLMGIIALVGLFTI
KVNSVTEFSGMAFVNGYSIMASVTPVHEIYFRQLGTMIIISGMDISPASSIIF
SNTLMEQOAGMSLVNVTYVNSYSLCIGKMAIYETQVNSGKLLNGYKRAQYLTGIG
LASLACMISGLVWESFTKGRARRAAETDCTVA
complement(7433..9046)
/gene="VANI"
/gene="VANI"
complement(7433..9046)
/note="YM8339.04C, VANI gene, len: 535, CAI: 0.16,
VANI_YEAST P23642 vanadate resistance protein"
/codon_start=1
/product="unknown"
/protein_id="CA89103.1"
/db_xref="GI:798885"
/db_xref="SWISS-PROT:P23642"
/translation="WGMFNLRSNIKKRAMDNGSLPISRNGSSNNIKDKRSEHNSNS

misc_feature

1.76 "overlap with lambda 7056"

CD5

/note="YM8339.02, unknown, len:1134, CAI: 0.14; P500017
1495..4899
/note="ATP/GTP-binding site motif A"
/codon_start=1
/product="unknown"
/protein_id="CA89101.1"
/db_xref="GI:798883"
/translation="MSNSKKKPVANVYVYRQOQDYNGMNAVGNPMVHPADVFNGAG
OYGPQHPAYVNSPLPNIPTPTPTAGALFPSHLMGSPVPSNMQSGYASAS
SLUKKATSRPVSNNNGNGNSNQNTNGAVTPSYVTRMGKNSRSNNNTNRYAT
NNHNGGQTRNNSGRRTEFARNNTEDDILPEMLDORPCINKVLPTDDAARTSLI
EWHNSDILSHIVKNFVKSNTLESAYLLEGKSDSKDVERKNLSLISFLTKGDCLN
FNNALIRSEKRTFLKSEALNFKVCVANTPKCLPTFESEALTENAEEDIDNGST
MNSALHHNIKAKDTRSLITIEFKSPKSDLPKLFQDLSRKNKRYLIESIDLVNT
DYPNSQEPENAVAYLPELINSMAIEYDLKRYKSLGSKPEYSLAPLYVSSRSVV
ANIYKGTSTHRLSVSVTAGNNDSNNNGNNSGTLTNNNSISGVSYGHSN
MLSTLSVSSVSLNEIDMLATKQCEVLDGYLEINRDVOTPIEBHSHLSVWKLS
KTTENSROFQDIPSPALPLENHEFMNDSQNGAIIIPOLLATPSPVPSNMQRPVLA
PPIPTQESQENFNVSAKVAASGSDIGRTYIGINIPRSKAEDICNVVCGILDSIT
LYPEKKIEVFTELEAPSAVOFYANGSFIDPVLKGNLKVCGKSGPLPKLSIATVNT
GASRVVYSLPEPAFKKEFHIDPOKTKHETLSLPDAEQLEDETSYGDIDQINLYSD
SHCCINIFNNISAIISLVEEMKEESTVONESGYTLKRAIEKFGGRKGLLINYGD
RCGNLNNKLIAKNGRFRYKVRPSYNTLRSLKEKRQNDIDDKAEFKPLMESEL
GSLDAHNDGGGETAGNTANGHEMESLEENENGEKNTGSGGLAVALSSDYKRAAT
SBETVEDYFNKSSGSDSDSDEYVIMSPDEPAALKSOTLRSSQOTVNSKRPVKI
DEEEVAYQMSDILNYSRSLROAPRAPPSLTLSNHSKNNETPMQDIFTEGFTANNKKRR
GSEFAHRTIRPGSDVAOYLAQVQSTFMYAANILGASAEQNTHPDE"
2845..2868
/note="P500017 ATP/GTP-binding site motif A"
5448..7076
/gene="ATRI"
5448..7076
/gene="ATRI"
/note="YM8339.03, ATRI gene (SNQ1), len: 542, CAI: 0.11,
ARI1_YEAST P13090 aminotiazole resistance protein"
conflict at C-terminus, integral membrane protein"
/codon_start=1
/product="Atr1p"
/protein_id="CA89102.1"
/db_xref="GI:798884"
/translation="MGNSQVLAETSEKVEYENETELPYKSSRDNNIGESLTAFAFQ
SDHEWVSNQKQNNPNFYKVAOEYLFETQMISOLLNAQOTPTLISNLTSSPSS
EENSQMSLMSPELVSGSESLISGRGIDYELKMLLVGYLYVIMSLIGITRYS
DTEFLTSARQGLGIAFLPNVLGTLGITYIGVFRKRTVYSGAPAAIGATIGCLFL
AGLIGTEDPKQMPAEVAVYSIAAFINFLVSYALPSTPINIHFSMDWIGSVYGVIG
LILLNENQAPISGMOQAYIIVILISIVLAVFYIYEIRFAKTPLLPRAVIDRHM
IOIMALEFGMGSGFGLFTFYEFQOLNIRQYALMAGGYEFLMGIIALVGLFTI
KVNSVTEFSGMAFVNGYSIMASVTPVHEIYFRQLGTMIIISGMDISPASSIIF
SNTLMEQOAGMSLVNVTYVNSYSLCIGKMAIYETQVNSGKLLNGYKRAQYLTGIG
LASLACMISGLVWESFTKGRARRAAETDCTVA
complement(7433..9046)
/gene="VANI"
/gene="VANI"
complement(7433..9046)
/note="YM8339.04C, VANI gene, len: 535, CAI: 0.16,
VANI_YEAST P23642 vanadate resistance protein"
/codon_start=1
/product="unknown"
/protein_id="CA89103.1"
/db_xref="GI:798885"
/db_xref="SWISS-PROT:P23642"
/translation="WGMFNLRSNIKKRAMDNGSLPISRNGSSNNIKDKRSEHNSNS

LKGRKVRPRSPKFOVLTSITLITAVLSLTLFSLFGSMGIGSTONGRSLSS
 SKSSNNYTTIDLEDEEYIDYDPEDIDPEVISKEPDGVOYHILISPGSEVLPRKDEEY
 QRELMLMDPDSIVEEYDLSNFEFGAPNGLETROHILLCLPLRAAADVILDMKRLMNLTY
 PHEILDIAFLVSDCSGSDGTLDALIAVSRHONCTLSIQIPEIDAVISDQKGDYKL
 LKYMGEQIINRYHOAFSPPEHNYDRPRSVOYIOKFGQVYIGGFSIDRAHAKVQGR
 RLKMGARANNTLANKALPKYHSMWYWRADAVYEGSVIODMSNYVYIYVNNWRPLP
 TPLGEOGYDLSNWMSEOLALAKTADDEDVYIGSVYAEQPTWYHAYIRDAEGDPN
 EAVDDIGGCVSILKAKIFRNGVOFPPEFTENHAEFGAKMKMGYRGGPLPHTT
 IWHIYEPDDDLKREIASERERKROSE
 complement(9295, 10827)
 /note="YM8339, 05c, unknown, len: 510, CAT: 0.13, possible
 coiled-coil"
 /codon_start=1
 /product="unknown"
 /protein_id="CAA89104.1"
 /db_xref="GI:798886"
 /db_xref="SWISS-PROT:003750"
 /translation="MTSTSSGSGTOSTYGVQLRNLMDLTREISHLDAVPELTRE
 VLFQNLNLTICSNFALDELVLNLTVOYDNGFRNLNLTLLRQSRQASQADKILLRE
 ENLDAPLSYLQGFQASEFTKSHSTVEYELKMSWSLALPHNEDEDELNNIEQONE
 INVLLPSPNLEKQIPSWLPNPPDHYKLPPEHNPITDLTKIKELIKESQSEKAA
 LNLKPSHSHTSASNPQPGLEDDEALIEQLIWMGALNEERKPTTEKSPFNENEE
 OYAKRYELNAREVYKTEFVNOLKTRKNEPILSEFLVLPESPOSHTKTIOLEOR
 .KSMTELPMLNLEPVYKTEKREKIRMAKERKPKSLKROEELISQPKRODQSGDELL
 NNEHAKDAADDTTPNALNNSTYINTRADEEDDINLFGISSSEDEENKSPAPEN
 LVAESEPTMTRADTTNTTPVAHNTNIDATVTSHPSTPENAPTSPADIAVDHDI
 TM"
 11297, 12043
 /gene="DAT1"
 11297, 12043
 /gene="DAT1"
 /note="YM8339, 06, DAT1 gene, len: 248, CAT: 0.09,
 DAT1_YEAST_P13483 oligo(dA)/oligo(dT)-binding protein"
 /codon_start=1
 /product="Dat1p"
 /protein_id="CAA89105.1"
 /db_xref="GI:798887"
 /db_xref="SWISS-PROT:P13483"
 /translation="MAKTLAOGRRPGSGRRKPGKGTLEGRKPPSGRRRRQDGGKET
 DSGDSQSRLLSRDMEAYALRELTSPSHSNHNSAAPPHMAASSTLPSLD
 YTHSGFMQOQQQQQQQQOQQOOLLQOQVYVPRKPTIHKILLSTGSGGHVNSNYNA
 DSHIHNSNHNHNSVNNMNPFTINGSQDDSSFLMGPIYVLDQRPYVLYDLSTS
 TNASQNPRTQPSAHITKNSDSTEKNTI"
 12315, 13205
 /note="YM8339, 07, unknown, len: 296, CAT: 0.15, previously
 noted as S22852 hypothetical protein 1 (Dat 3' region)"
 /codon_start=1
 /product="unknown"
 /protein_id="CAA89106.1"
 /db_xref="GI:798888"
 /db_xref="SWISS-PROT:P46963"
 /translation="MDSLRLQFLQIYMKNLQTLRLKTRDSTSSSTTPPSSQOKLN
 NDIQFYLRNRYHHEDEPHOCLPFTOMMDPLDLVYVYRILRYPHASHSNPY
 TKVLEVLMDIDLVELCPQDMKSLTNQATKEFLDLSKILIHIDATSVHTPSD
 TPLDATWYVKTEERTKYDKESLQRTESLLKRDKLKLAFOQFNSDTAIPDLQ
 TQPTANLILHMEADRELKRSKETSYLERPNDILDESEKSLMTTHTETDSDGP
 KDDYKIRALNDIAASYIY"
 14194, 16956
 /note="YM8339, 08, unknown, len: 920, CAT: 0.15"
 /codon_start=1
 /product="unknown"
 /protein_id="CAA89107.1"
 /db_xref="GI:798889"
 /db_xref="SWISS-PROT:003758"
 /translation="MTPEFTSSSRKNGNPPKLSYSTDENIHLKRRQOQSSNSTDN
 SLHPSSGQTPRRSDODDTSASNTNLNDRLOREESLSLMDCTQSRLSHRAMVLVD
 VLPSSENNALHRLIPQGNVDPDRHDEPPSVQEVRTQMTLPNSDSVRSQLTAVP
 GSENCNNAHTSLNLHPLOTHLQTLNTSTSGGQSHSSSDTNISQIPEDDINSQD
 NFIQDLTKPLSLPEIDIRIKTASIPHEREOSILKEYISGIIIGYCLIEAR
 SSQPLKFMFVYVTLTAVTSVIDROKRTKRLRLRWDLSASWYIIPRSGIYVY
 GERDDDAIITGLSNRELKPTNTKYKKFMRKLPOLLDYTCQDQPSHCLLPSPGID
 KTKNNCKSGIKGVSGHIGTKGSPITLMDADNDSINTYITDILYGVKGRFTSK
 LTKNNKRYNRYVMPFAGVGTNOONETCTQLRQLNKSNESLIEFEFLINKIFKLELN

EAISNVDIHDITDGLDNEDELDSEILRRKLDOLHINRIIDDTASOSPYSKNNMA
PKENLVETELRKFKKNNKNSLSFHSHESSSSSTPHVNSGLIYIVSKKPOS
TLPYSPRLRKTKKFKPSKSEKNNQRLMGMLPEGYKPLTLKDLHITCISNNKA
GKRPESVITTEVITAKSDNSIPKFCLEMLNEKRLKTLKFLYOKVHEYR
KKEENNAKINELYNRRDHEFPKELFTNFIISOJINNDISLAGLVNIIDLDHDIRK

Query Match 4.9%; Score 143.8; DB 8; Length 27559;
Best Local Similarity 50.5%; Pred. No. 1.2e-17;
Matches 463; Conservative 0; Mismatches 412; Indels 42; Gaps 3;

QY 1 tgaactgttgaacaaagaagctgacaaagcattgtaacagatattac 60
Db 16332 TGAATGTTGATGAGTGAAGTAAACAACTAAAGCAAAATTTTTCAGTATCA 16391
QY 61 ttcaaaatttgtaataatgaacaaagttcaacaaagatttctgtaataatgaagt 120
Db 16392 GAAAAAGTTCATGAAATTCGTAATAAATTTGAAGAAAACATGCGAAGCTAAAGAGACT 16451
QY 121 atataatgaataaggaagacgtaaggccaaagaaactgaatttcaagatttattac 180
Db 16452 TTATATATAGAAATAGAGATCAATTCATCTCCAAAGAGCTTTTGTTCACAAATTTTATATC 16511
QY 181 ttcaacagctgttcaacgatacgaagaacattgcaactgaaagttagtgtaacact 240
Db 16512 TGATCAATTAATAACATGATATTGACAGCTTTAGCTGATTAAAGTTAATATATGATTT 16571
QY 241 atcaacatttttaaaacaggtgtaacccataaacaacactaaagcagcattac 300
Db 16572 ACATGATATCTTCATAAAGCAAAATACATATCTTTGAAGAGAAAACGAGGATATTATTTTC 16631
QY 301 tgaagatcaatatacgcacacaggaagtaagctgacgacgacgacgacgacgac 360
Db 16632 GAAGAAAGAGAGGTTTC-----CAATCCTCCTCCCGCATCATCAATTAATATTT 16692
QY 361 aacgcagtaacattcattccaaagaagtagttatttaccagcggaagcagcagc 420
Db 16693 CTTTACAAACCAACATTT-----TACCAATGCTGATCACCAC 16715
QY 421 ttccctgaatttcaagcagatgttcataaaggttgagttgagtccttccatca 480
Db 16716 CGCTACTAAGTTTACTCGCAAAATGTGATGAGGAAAAAGTTAAGCCATTTCACATA 16775
QY 481 caaacgagacatlaactgtaactgtaacttaataagacataagaaacttaattcc 540
Db 16776 TAAGAGGAGATGTTACAGTTAATTTAGCTCAACCCCAACATTAAGAAAGCTTATGTCGC 16835
QY 541 aagtttgaagcgtcctaagttgtagtttattgctgcgagtaagtaataattga 600
Db 16836 CAATCTTGAAACGTGTTGTGTGAGATTTTATGTGTGAGTAAGCAATCAATTTGA 16895
QY 601 aaacacattgagcgaagatgatataccatttctgttagcgaagtgacaata 660
Db 16896 TAACCACTTGGGTTCTATGAAAGTGAATTCGGTGTGATGAGAAAAGTTAACAATTTA 16955
QY 661 aaaaacattgaaaaaattcgttacttcttatalagataatagatattgagttc 720
Db 16956 GACCATTAACAACACTAACGATATCTCGCAGTTTTCATAAATATATATATAAAGCAGACTA 17015
QY 721 gctatagatgaagattatcgcgccttctgtatccattatlaataaattcttc 780
Db 17016 TTTTCTT-----AACCATTTAATATACCTATTAATTAATTAATTAATTAATTAAT 17069
QY 781 taaatgatatttctgagccttctgttctgtctgtatttcttcttcttcttcttcttct 840
Db 17070 TATGATGACGTCATTAATTTCTTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCT 17129
QY 841 gatgaaaaacctgattatattatattatattatattatattatattatattatattat 900
Db 17130 TAGTATATACGAATTTCTGCACATCAGCAAAATATATGCTATATCATATGCGACAGTAT 17189
QY 901 caggacaaaacacataga 917
Db 17190 TACTTAAGAGAGAGAGA 17206

RESULT 11
SCYNR068C/c
LOCUS SCYNR068C 2023 bp DNA PLN 11-AUG-1997
DEFINITION S.cerevisiae chromosome XIV reading frame ORF YNR068C.
ACCESSION Z71683.Y13139
VERSION Z71683.1 GI:1302599
KEYWORDS
SOURCE baker's yeast.
ORGANISM Saccharomyces cerevisiae
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetaceae; Saccharomycetaceae; Saccharomycetes.
REFERENCE 1 (bases 1 to 2023)
Duesterhoef, A., Floeth, M., Frit, C., Heuss-Neltzel, D., Hilbert, H.
and Moestl, D.
Unpublished
2 (bases 1 to 2023)
MIPS.
Direct Submission
Submitted (29-APR-1996) Data collected by MIPS on behalf of the
European yeast chromosome XIV sequencing project. MIPS at the
Max-Planck-Institut fuer Biochemie, Am Klopfersplitz 18a D-82152
Martinsried, FRG; E-mail: Mewes@mips.emblnet.org
FEATURES
location/Qualifiers
1..2023
/organism="Saccharomyces cerevisiae"
/db_xref="taxon:4932"
/chromosome="XIV"
complement(968..1786)
/note="ORF YNR068C"
/codon_start=1
/protein_id="CA96350.1"
/db_xref="GI:1302600"
/db_xref="SWISS-PROT:P53754"
/translation="MRNRONK NMTNINLGLSVEEREKLENI RLELYCOMAANSIPHD
PPEISSLETETICTTNTKDC PVPFSDLLKHKYNEIKFKILEIENTEAYROER
TKNOTKINLLADPARASLRPSIDPFDIMSSYTKDYQVIANEAVYVAKLTK
LVGRKSPVASSPISSTIPRISNRKKTSPSHSHSVSHRSKNEMNOVSKRTLL
LNIRYNDKFATVIFSESCISRSYFLRVKLHPDKGVSAEIDIPQVKNSTI"
BASE COUNT 597 a 361 c 352 g 713 t
ORIGIN
Query Match 2.5%; Score 75; DB 8; Length 2023;
Best Local Similarity 55.6%; Pred. No. 0.00015;
Matches 144; Conservative 0; Mismatches 115; Indels 0; Gaps 0;
QY 430 attcagaccagatgttcataaaggttaggaattgctccttaccagtaacagaga 489
Db 1199 ACTTTCACATCGGAATCAATGAATGGAATAGTCACTTAAGTATAGAGAAC 1140
QY 490 catlaigtgaactggaatttaataagacataaagaaacttaattccaagtttga 549
Db 1139 ACTGTTATGAAATATAAATATATAGATGATTTAAGAGTACCAATGAGCCAACTTTGA 1080
QY 550 aagcgtcctaagtgttaggttattatgtgttcgagtaagtaattgaaccacatc 609
Db 1079 AAGTGTATTAGCTCGAGAGCATATATTTCTCGCGCTAAACCTCATTTTGAATAAGGTGT 1020
QY 610 tggcgtagcgaagattgatalcccatcttctgttagcgaagtgacaaaataaataacat 669
Db 1019 TGGATCTGCTGAATGATGATATCCAGTTCAAGTTAAAAAGCTCTTTATTTATGATGATGCG 960
QY 670 tagaaaaattctcgctac 688
Db 959 TGTTAATAAGATGTTTC 941
RESULT 12
SCU32938 4775 bp DNA PLN 03-DEC-1996
LOCUS SCU32938
DEFINITION Saccharomyces cerevisiae zds2p (ZDS2) gene, complete cds.

ACCESSION VERSION U32938
KEYWORDS U32938.1 GI:1236326
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
FEATURES
source

baker's yeast.
Saccharomyces cerevisiae
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; Saccharomycetaceae; Saccharomyces.
1 (bases 1 to 4775)
Yu, Y., Jiang, Y.W., Wellinger, R.J., Carlson, K., Roberts, J.M. and
Stillman, D.J.
Mutations in the homologous ZDS1 and ZDS2 genes affect cell cycle
progression
Mol. Cell. Biol. 16 (10), 5254-5263 (1996)
96413276
2 (bases 1 to 4775)
Stillman, D.J.
Direct Submission
Submitted (01-AUG-1995) David J. Stillman, Division of Molecular
Biology and Genetics, Department of Oncological Sciences,
University of Utah Health Sciences Center, Salt Lake City, UT
84132, USA
On Mar 27, 1996 this sequence version replaced gi:387502.
Location/Qualifiers
1..4775

[illegible]

| | Query Match | 2.48; | Score 71.8; | DB 8; | Length 4775; |
|---------|---|--------------|-------------|----------------|----------------|
| | Best Local Similarity | 71.8%; | Pred. | No. 0.00059; | |
| Matches | 94; | Conservative | 0; | Mismatches 37; | Indels 0; Gaps |
| Oy | 2401 gataaacggaagctcgaagtlttgattgcctcgacacgtcccttgcacaatgaaalccggcagc | 2461 | | | |
| Db | 804 GAGAAAGGAGAACTGCACGTTAATTCGCCCAACTCTTTGGATTGGAATAATCAGAAT | 863 | | | |
| Oy | 2461 gttaaagaacctaaaagaatttcgatattgggtcaaatgattttctaactatgatccagaatta | 2521 | | | |
| Db | 864 GTGAAAATTGAAGCATTTGTGATTTGGTTCANATGAAATTCACCCTAGACTA | 923 | | | |

| | | | |
|----|------|-------------|------|
| QY | 2521 | gataataaatt | 2531 |
| | | | |
| Db | 924 | GAGTCAAAAGT | 934 |

| RESULT | 13 | | | | |
|------------|---|--------|-----|----|-------------|
| CNS06KB/C | | | | | |
| LOCUS | CNS06KB | 780 bp | DNA | SR | 10-JAN-2001 |
| DEFINITION | T7 end of clone AT0A002H08 of library AT0A from strain CBS 4311 | | | | |
| DEFINITION | of <i>Saccharomyces servazzii</i> , sequence tagged site. | | | | |

| | |
|-----------|-------------------------|
| ACCESSION | AL402561 |
| VERSION | AL402561.1 |
| KEYWORDS | GI:12161691 |
| SOURCE | STS. |
| | Saccharomyces servazzii |

ORGANISM *Saccharomyces servazzii*
- Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes
- Saccharomycetales; Saccharomycetaceae; Saccharomyces.
REFERENCE
1 (bases 1 to 780)
Casaregola, S., LePingle, A., Bon, E., Neuveglise, C., Nguyen, H.,

TITLE
Artiguenave, F., Wincker, P. and Galliard, C.
Genomic Exploration of the Hemiascomycetous Yeasts: 7.
Saccharomyces servazii
JOURNAL
FEBS Lett. 487 (1), 47-51 (2000)
PUBMED
11152882

| | |
|-----------|---|
| REFERENCE | 2 (bases 1 to 780) |
| AUTHORS | Souciet,J.L., Aigle,M., Artiguenave,F., Blandin,G., |

TITLE
Genomic Exploration of the Hemiascomycetous Yeasts: 1. A set of yeast species for molecular evolution studies(1)
JOURNAL
FEBS Lett. 487 (1), 3-12 (2000)

PUBMED 11152876
 REFERENCE 3 (bases 1 to 780)
 AUTHORS Genoscope.
 TITLE Direct Submission
 JOURNAL Submitted (07-SEP-2000) Genoscope - Centre National de Sequençage,

COMMENT

2 rue Gaston Cremlieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail : seger@genoscope.cns.fr - Web : www.genoscope.cns.fr)

This STS is part of a random genomic sequencing program of thirteen yeast species: *Saccharomyces bayanus* var. *uvatum*, *Saccharomyces exiguus*, *Saccharomyces servazzii*, *Zygosaccharomyces rouxii*, *Saccharomyces kluyveri*, *Kluyveromyces thermotolerans*, *Kluyveromyces lactis* var. *lactis*, *Kluyveromyces marxianus* var. *marxianus*, *Pichia angusta*, *Debaryomyces hansenii* var. *hansenii*, *Pichia sorbitophila*, *Candida tropicalis* and *Yarrowia lipolytica*. Genomic inserts of 3 to 5 kb were prepared and both extremities were sequenced. See keywords for description of this sequence and for the sequence of the other extremity of this insert.

| FEATURES | Location/Qualifiers |
|----------|---------------------|
| source | 1. .780 |

```
misc.feature
complement(<256, >780)
/note="T7"
/clone_lib="AT00A"
/db_xref="taxon:27293"
/strain="CBS_4311"
/organism="Saccharomyces servazzii"
```

```
misc_feature
    /note="similar to Saccharomyces cerevisiae ORF YMR275c
    BUL1; ubiquitination pathway protein]"
    /evidence="not_experimental"
    complement(<265..>771)
```

| BASE COUNT | ORIGIN |
|------------|-----------------------------|
| 251 a | 110 c 138 g 255 t 26 others |

| | | | | |
|-----------------------|-------|-----------|--------|------------|
| Query Match | 2.3% | Score 68: | DB 11: | Length 780 |
| Best Local Similarity | 56.2% | Pred. No. | 0.0032 | |

| | | | | |
|------------|---|--|---|--------|
| | D | 172229 | TTCTTCACATTAATTATATATATATTTTAAATATTTTATATATATATATATATTTTTTTAA | 172286 |
| OY | 526 | ggaaacttaattcccaagctttggaagcgcgatgtlgtlaagllttaatcgltcgaagt | 585 | |
| D | 172289 | AATAATTAATAATTAATTAATGCTGTATTAAAAATTTATATATATATATATATCATATAT | 172348 | |
| OY | 586 | aatgatctaattbgaanaacatcttcgtagagaagatgataccccatttcctgttag | 645 | |
| D | 172349 | AATTATATTTATATATAAAAAAAGAAATAAACCTGTGAATTAATTCGATATATTTCTTAATTATTA | 172408 | |
| OY | 646 | gcaagtcacaataaaaaaacatlagaaaaaatcccgccttaatttcctcatagataag | 705 | |
| D | 172409 | ATFAAATTAATAATATAAAAATAATAATAATAATAAAAATATTAAGTACATATATACATATC | 172468 | |
| OY | 706 | atatatgtagtgcttgccttagatagaaggatlattalcgcgtcccttgttatcccatla | 765 | |
| D | 172469 | ATATAAATATATACATATATATATATATGCATATTTATATATATATATATATATATATATAT | 172528 | |
| OY | 766 | ttaataaatcttcttaaagaatcttcgcgtgcctcttgctgtgcctcgtatcttttt | 825 | |
| D | 172529 | TATATATATATATATATAGTTATGCTTATGATATTAATTAATCTTTTGATTTTATTTTATTT | 172588 | |
| OY | 826 | tttttggaccactgtagtgaaaaaccttgcattgatattcatcttaatttaagttaact | 885 | |
| D | 172589 | TTTTTCAATATGTAAACGGTAGTGTCCTTTGATGATTAATTTCCAAAATAATAAGTATAT | 172648 | |
| OY | 886 | aaatatcagagat-ctcaagaaacaaacatalagaatttcttctgtcagaagaaataaacg | 944 | |
| D | 172649 | TTAAAAATTTTATTTATTTACTCCCAATAATAATATATATTTTATCTTTAAGAAATATA | 172708 | |
| OY | 945 | aaataaattgat | 956 | |
| D | 172709 | AAAAATATTTAT | 172720 | |
| RESULT | 15 | | | |
| LOCUS | AC005140 | | | |
| DEFINITION | Plasmodium falciparum chromosome 12 clone 3D7, *** SEQUENCING IN PROGRESS ***, 4 unordered pieces. | | | |
| ACCESSION | AC005140 | | | |
| VERSION | AC005140.8 | | | |
| KEYWORDS | HTG; HTGS; PHASEI. | | | |
| SOURCE | malaria parasite P. falciparum. | | | |
| ORGANISM | Plasmodium falciparum | | | |
| REFERENCE | Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium. | | | |
| AUTHORS | 1 (bases 1 to 310779) | | | |
| REVIEWERS | Hyman,R.W., Fung,E.L., Qin,F., Rowley,D., Mao,J., Tamaki,T., Kirdi,O.B., Conway,A.B. and Davis,R.W. | | | |
| JOURNAL | Plasmodium falciparum 3D7 chromosome 12 | | | |
| REFERENCE | Unpublished | | | |
| AUTHORS | 2 (bases 1 to 310779) | | | |
| TITLE | Hyman,R.W., Qin,F., Fung,E.L., Conway,A.B. and Davis,R.W. | | | |
| JOURNAL | Direct Submission | | | |
| COMMENT | Submitted (18-JUN-1998) Stanford DNA Sequencing and Technology Center, Stanford University, 855 California Avenue, Palo Alto, CA 94304, USA | | | |
| | On Aug 12, 2000 this sequence version replaced gi:8810455. | | | |
| | * NOTE: This is a 'working draft' sequence. It currently consists of 4 contigs. The true order of the pieces * is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. * This record will be updated with the finished sequence * as soon as it is available and the accession number will | | | |

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 12, 2002, 10:38:12 ; Search time 4622.84 Seconds

(without alignments)
6512.739 Million cell updates/sec

Title: US-09-165-460a-1

Perfect score: 1825

Sequence: 1 acctaccctttttctatct.....aaaataaacgattaacatt 1825

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : GenEmbl:*

1: gb.ba:*

2: gb.htg:*

3: gb.in:*

4: gb.om:*

5: gb.ov:*

6: gb.pat:*

7: gb.ph:*

8: gb.pl:*

9: gb.pr:*

10: gb.ro:*

11: gb.sts:*

12: gb.sy:*

13: gb.un:*

14: gb.vi:*

15: em.da:*

16: em.fun:*

17: em.hum:*

18: em.in:*

19: em.om:*

20: em.ov:*

21: em.ov:*

22: em.pat:*

23: em.ph:*

24: em.pl:*

25: em.ro:*

26: em.sts:*

27: em.sy:*

28: em.un:*

29: em.vi:*

30: em.htgo.hum:*

31: em.htgo.inv:*

32: em.htgo.rod:*

33: em.htg.hum:*

34: em.htg.inv:*

35: em.htg.rod:*

36: em.htg.other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB | ID | Description |
|------------|--------|-------------|--------|----|-----------|---------------------|
| 1 | 1823.4 | 99.9 | 1825 | 8 | SCYR117W | 249617 S. cerevisia |
| 2 | 1553.4 | 85.1 | 1706 | 8 | SCU77137 | 077137 Saccharomyc |
| 3 | 342 | 18.7 | 1555 | 8 | SCYR116W | 249616 S. cerevisia |
| 4 | 195 | 10.7 | 99360 | 8 | NCB11N2 | AL513444 Neurospor |
| 5 | 187 | 10.2 | 36920 | 8 | SPAC3H1 | 268144 S. pombe ch |
| 6 | 137.2 | 7.5 | 1275 | 8 | AF353722 | AF353722 Arabidops |
| 7 | 136 | 7.5 | 686 | 6 | AB6270 | AB6270 Sequence 92 |
| 8 | 136 | 7.5 | 686 | 6 | ARI55763 | ARI55763 Sequence |
| 9 | 121 | 6.6 | 935 | 8 | SCYR118C | 249618 S. cerevisia |
| 10 | 117.4 | 6.4 | 1859 | 9 | AF064867 | AF064867 Homo sapi |
| 11 | 117.4 | 6.4 | 2963 | 9 | AB016068 | AB016068 Homo sapi |
| 12 | 117.4 | 6.4 | 2966 | 9 | HSY13834 | Y13834 Homo sapien |
| 13 | 117.4 | 6.4 | 2968 | 6 | E32056 | E32056 Human AF1. |
| 14 | 114.2 | 6.3 | 1340 | 3 | AF358443 | AF358443 Physarum |
| 15 | 114.2 | 5.4 | 78592 | 3 | AC004335 | AC004335 Drosophi |
| 16 | 98.4 | 5.4 | 102413 | 2 | AC020024 | AC020024 Drosophi |
| 17 | 92.8 | 5.1 | 41327 | 3 | CBRG47J19 | AC084676 Caenorhab |
| 18 | 91.6 | 4.8 | 40767 | 3 | CEC04F12 | Z81461 Caenorhabdi |
| 19 | 88 | 4.7 | 582 | 8 | CNS018P | AL10644 Botrytis |
| 20 | 86.6 | 4.5 | 660 | 8 | CNS01AH3 | AL112927 Botrytis |
| 21 | 81.6 | 4.0 | 2161 | 9 | AK027874 | AK027874 Homo sapi |
| 22 | 72.2 | 3.7 | 1072 | 6 | AX053334 | AX053334 Sequence |
| 23 | 66.6 | 3.6 | 124808 | 2 | AP000652 | AP000652 Homo sapi |
| 24 | 66.6 | 3.6 | 16488 | 1 | AE004992 | AE004992 Halobacte |
| 25 | 65.6 | 3.6 | 256172 | 2 | AC005139 | AC005139 Plasmodiu |
| 26 | 65 | 3.5 | 169423 | 2 | AC069309 | AC069309 Mus muscu |
| 27 | 64.4 | 3.5 | 18306 | 2 | AC083946 | AC083946 Mus muscu |
| 28 | 64.4 | 3.5 | 1245 | 6 | AX055202 | AX055202 Sequence |
| 29 | 63.6 | 3.5 | 19715 | 1 | AE000555 | AE000555 Helicobac |
| 30 | 63.6 | 3.5 | 234112 | 3 | PFMAL4P2 | AL035475 Plasmodiu |
| 31 | 63.6 | 3.4 | 13383 | 4 | AE001374 | AE001374 Plasmodiu |
| 32 | 62.4 | 3.4 | 98734 | 2 | PFMAL1P2 | AL031745 Plasmodiu |
| 33 | 61.8 | 3.4 | 14943 | 1 | AE001528 | AE001528 Helicobac |
| 34 | 61.6 | 3.3 | 12029 | 3 | AE001392 | AE001392 Plasmodiu |
| 35 | 59.8 | 3.3 | 110000 | 2 | AC087899 | AC087899 Mus muscu |
| 36 | 59.8 | 3.2 | 114736 | 3 | PFMAL3P3 | 298547 Plasmodium |
| 37 | 59.2 | 3.2 | 253305 | 3 | PFMAL3P7 | AL034559 Plasmodiu |
| 38 | 58.8 | 3.2 | 98899 | 2 | AL603743 | AL603743 Danio rer |
| 39 | 58.8 | 3.2 | 318221 | 2 | PFMAL13P3 | AL049184 Plasmodiu |
| 40 | 58.8 | 3.2 | 318503 | 2 | AC074166 | AC074166 Mus muscu |
| 41 | 58.4 | 3.2 | 172239 | 2 | AC069587 | AC069587 Homo sapi |
| 42 | 58.4 | 3.2 | 12029 | 3 | AE001412 | AE001412 Plasmodiu |
| 43 | 58.2 | 3.2 | 187469 | 9 | AL357312 | AL357312 Human DNA |
| 44 | | | | | | |
| 45 | | | | | | |

ALIGNMENTS

RESULT 1

SCYR117W 1825 bp DNA 11-AUG-1997

LOCUS SCYR117W

DEFINITION S. cerevisiae chromosome X reading frame ORF YCR117W.

ACCESSION 249617 Y13136

VERSION 249617.1 GI:1015836

KEYWORDS

SOURCE

ORGANISM

REFERENCE

1 (bases 1 to 1825)

REFERENCE

AUTHORS Rose, M., Koetter, P. and Ertlan, K.D.

JOURNAL Unpublished

2 (bases 1 to 1825)

REFERENCE

AUTHORS

JOURNAL

Submitted (25-SEP-1995) Data collected by MIPS on behalf of the European yeast chromosome X sequencing project. MIPS at the Max-Planck-Institut fuer Biochemie, Am Klopferspitz 18a D-82152

| Db | 1661 | GTGAGTGAAGAAGAAAGAAACCTATCTATAGAGTACACATATTAGCATGTACCGTTAAAT | 1740 |
|----|------|---|------|
| Oy | 1741 | tcacgttcggtatgctcatctcttcaactacatacaacagcgctactctactataagaataaaga | 1800 |
| Db | 1741 | TCACGTTGCGTATGTCTATCTATCTACACATCACTCACAGTATCTACTATATAGATATAAAGA | 1800 |
| Oy | 1801 | aagaaaaataaacgalttaaacalt | 1825 |
| Db | 1801 | AAGAAAAAATTAACGATTAAACATT | 1825 |

| RESULT | 2 |
|------------|--|
| LOCUS | SCU77137 |
| DEFINITION | SCU77137 1706 bp DNA |
| ACCESSION | Saccharomyces cerevisiae zinc metallo-protease (STE24) gene, |
| VERSION | complete cds. |
| KEYWORDS | 077137 077137 |
| SOURCE | 077137.1 GI:1679740 |
| ORGANISM | Baker's yeast. |
| | Saccharomyces cerevisiae |

| REFERENCE | AUTHORS | TITLE |
|---------------------|--|--|
| 1 (bases 1 to 1706) | Fujimura-Kamada, K., Nouvet, F. J. and Michaelis, S. | A novel membrane-associated metalloprotease, Stc24p, is required |

JOURNAL
REFERENCE
AUTHORS

J. Cell Biol. (1996) In press
2 (bases 1 to 1706)
Michaellis, S. and Fujimura-Kamada, K.

JOURNAL Submitted (01-NOV-1996) Cell Biology and Anatomy, Johns Hopkins University School of Medicine, 725 North Wolfe Street, Baltimore, MD 21205, USA

| FEATURES | Location/Qualifiers |
|----------|---------------------|
| source | 1. .1706 |

```

/organism="Saccharomyces cerevisiae"
/db_xref="taxon:4932"
/chromosome="X"
/map="YJRL17W"
73..1434
gene
/feature="STE24"
73..1434
CDS

```

```

BASE COUNT      494 a      366 c      275 g      571 t
ORIGIN
/gene="STE24"
/function="first step of N-terminal processing of the
yeast a-factor precursor"
/Note="Ste24p"
/codon_start=1
/product="zinc metallo-protease"
/protein_id="AAB38271.1"
/db_xref="GI:1679741"
/translation="MEDKILIDHPNIPMKLIISGSIAQSFSESYITVROYKSL
KLPVLEDEIDDETFHKSRRYSRAKAKESIFGVYVINAQKLVITKIDLPKIHMM
LNAVLVPRFHMVSTVAQSLCFGLGLSLSDVDLPISYSHVLEQGEKFNLYT
WIDMIRKSLTALAIGDPIYLFLKIDFSDPTDPLWYIMVLELVQOILMTIIPV
PMNRKFPPLDEDGELKSIIDSLADRGVDPDLKIVIDGSKRSSNAPFGLPTSS
VLEPDTLVNSSTDDEITAVLAHEIGHOMKNIVMVVLESLHTEFLSFLSIYRNN
YVNEGFLEKSGSFPVDVPTTKKEPPIIGFMPLPNDLLTFPECAOMQVMSLISRTHH
ADAYAKKLTGKQNCRALIDLOIKNLSTNANVDPVLYSSHYHSHTPLAERLTALDYV
KKN

```

| | | | | | |
|----|----------------------------|---|------------------|-----------|--------------|
| | Query Match | 85.1% | Score 1553.4 | DB#: | Length 1706; |
| | Best Local Similarity | 99.9% | Pred. 0.23e-306; | | |
| | Matches 1554; Conservative | 0; | Mismatches 1; | Indels 0; | Gaps |
| Ox | 271 | aaaccattctataattacaggaagaagaaaaaggaggaatatagaactgc | 330 | | |
| Dd | 1 | AAACCACTTCATTAATTACAGGAAAACGAGGAAAAAAGAGGAATAATGAAAACGTGC | 60 | | |

| | | | | | | | | | | |
|----|------|----------------|-----------------|------------------|------------------|----------------------|---------|-----------|----------|-----|
| QY | 331 | aggccttatc | taagtttg | atcctaaga | agatcttc | tgacac | cttaata | tatcccg | tggaaa | 390 |
| Db | 61 | AGGCGTTTATTCAG | TTTGATCTTTAAG | ACGATTTCTGCAC | CACTCTTAATATC | CCGTGGAA | | | | 120 |
| QY | 391 | ttaacatcttc | tgagtlc | tgatctgc | ttgccaaatcttc | cttgaa | ctcctga | cttcga | ctgacga | 450 |
| Db | 121 | TTAACTATTTCTGG | GTTCGATTTGCC | CAATTTCTTG | CAATCTTA | CTTGACGTACAG | | | | 180 |
| QY | 451 | cagtaacga | agactctc | tgaaacaa | agttctgc | acactgtgc | tgbaa | gcgaactga | ttgat | 510 |
| Db | 181 | CAGTACGACGAAG | CTATCTGAAACAAAG | TGTGCACTGTGTG | GTGGAAGACGAAT | TGATGAT | | | | 240 |
| QY | 511 | gaacatcttc | alcaaa | lcaagg | aactcc | ggcgga | agccaa | gblctcc | atcttc | 570 |
| Db | 241 | GAAACTTTTCAAT | CAATCAAG | CAACTAC | CCGGGCAAGGCAAG | TTCTCATTTTCGGT | | | | 300 |
| QY | 571 | gacgccta | taactag | cccaaa | actagttt | taacaa | tagaac | ctctcc | taaatc | 630 |
| Db | 301 | GACGCTATTAAC | CTAAGCCCCAAAG | CTAGTTTCA | CAAAATGCAAC | CTTCCCTTAATC | | | | 360 |
| QY | 631 | tggcaca | tgagccgt | ttctctta | ttgaatgc | agtcgc | agtcga | ttcata | tgatctcc | 690 |
| Db | 361 | TGGCACAATGGCG | GTTCCTTTATGA | ATGCAATCCTG | CGAGTCAATTTCA | TATGATGCTCC | | | | 420 |
| QY | 691 | actgtcga | caga | glttat | gtcttc | tggtgc | ttctat | ccagtlgt | ctacctgt | 750 |
| Db | 421 | ACTGTGCA | CAGAGTTTATG | CTTGTGGGTCTTAT | CCAGTTGTGCTAC | CTTGGTTGAT | | | | 480 |
| QY | 751 | ttggacact | ctctta | agcatctt | gtctctgc | tgaaaaaa | atttg | tttca | taaatg | 810 |
| Db | 481 | TTGGCACTCTCT | CTAATATAG | CAATTTGTCTG | GAAGAAAAATTTG | ATTTCATTAATAATG | | | | 540 |
| QY | 811 | accgtc | caactat | tgatca | ccagatga | lcaaa | gagtc | tgact | gtg | 870 |
| Db | 541 | ACCGTCAAC | TAATGATATC | CCGATATGATCA | AGACTGACTG | ATTGGGATATG | CTATTCG | | | 600 |
| QY | 871 | ggccca | atccct | taactgt | ttccctaa | gactt | tgataa | atccct | actga | 930 |
| Db | 601 | GGCCCAATC | CTTTACCTGTCTT | TAAGATCTTTATTAAT | TCCCTCACTG | ATTTCTCTTGG | | | | 660 |
| QY | 931 | tacatata | ggtcttc | gtgtgt | gtgc | aaatct | tagca | atgc | atcatct | 990 |
| Db | 661 | TACATTATG | GTCTTCTTGTGTG | TGCCAAATCTTAC | CCATGACAAATCAAT | CTTCAAGTCTTC | | | | 720 |
| QY | 991 | atcatgc | caactgt | ttaa | taagttca | ctccatct | gtgaga | gcgtga | ac | 105 |
| Db | 721 | ATCATGCCCAT | GTTTAATAA | GTTCAC | TCTCATTTGGAGGAC | GTGACATGAAAAATCTATT | | | | 780 |
| QY | 1051 | gaag | tttgc | ccga | laga | gttgg | ttccct | laga | taaga | 111 |
| Db | 781 | GAAATTTTGGCG | CGATAGAG | TGTGGTTCCTTAT | AGATTTTGTTCAT | TGACGGCTCA | | | | 840 |
| QY | 1111 | aaa | gaatcttc | atctca | aaagcatatct | taacag | gttgc | atctc | acctca | 117 |
| Db | 841 | AAAAATCTTCT | CAATCAATAA | GCATATTTCA | CAGGTTTGCAT | TCACTCAAGAGAAAT | | | | 900 |
| QY | 1171 | gtttgtc | gcga | acttgc | tgaa | caagta | atctgc | atga | aatctga | 123 |
| Db | 901 | GTTTGTTC | GCACACTTTAT | GTATGATACAG | TATTTCTATGATGAAT | TACGGCTGTTTTGGC | | | | 960 |
| QY | 1231 | catga | aatc | ggtlca | ctgcgc | aaaaaa | ccacat | cgltta | atay | 129 |
| Db | 961 | CATGAATTC | GGTCTAC | TGSCAAAAA | ACACATCTGTATAT | TATGTCATCTTTACTCAATG | | | | 102 |
| QY | 1291 | caacact | ctccatcttc | ctccctt | taaccca | gatactca | agaa | ataca | taatttca | 135 |
| Db | 1021 | CACACCTTCT | CAATTTCTCC | TTTTAC | CGCATCTAC | AGAAATATCATTTTCAAC | | | | 108 |
| QY | 1351 | accctgc | gcttcttc | taaga | aaatc | caactgc | gagctt | ttgtga | cccg | 141 |
| Db | 1081 | ACCTTGGCG | CTTTTCTTA | AGAAAGTCC | ACGTGGCA | GTTTGTGTATCCCGTAT | CACTAAG | | | 114 |
| QY | 1411 | gaatcc | ccat | atcat | tgatlt | atltat | ltaag | actat | taactca | 147 |

```

Db 1141 GAATTCGCCATATGATGATTGATTATTAAGCACTATTAACCTCCACTCGAATGT 1200
OY 1471 gccatgcaattcgtatgaattatcttcagaactcagaatatacaagctatctat 1530
Db 1201 GCCATGCAATTTGTATGTATGTTAAATTTCCAGAACTCAATGAATATCAACTGATCTTAT 1260
OY 1531 gctaaataatggctacagaacaaatctatgtaaggctcattatgaltacaaatcaaa 1590
Db 1261 GGTAAATAATTTGGCTACAGACAAAATCTATGTAGGGCTCTAATGATCTACAAATCAAA 1320
OY 1591 aaccttccacatgaatgtagatccctcgtatcttaagctatcatatccatccaact 1650
Db 1321 AACCTTCCACCATGAATGTAGATCTCTGTATCTAGCTATCATTTATCCCATCAACT 1380
OY 1651 caagctgaaagaatgcagccctctagaactatgtagtgaataaagaataaactata 1710
Db 1381 CTAGCTGAAGAATGTAGCCCTCTAGACTGTATGTAGTGAAGAAAGAAAATTAATCTATA 1440
OY 1711 gagtacacatatagcatgtagcgttaaatcaagcttcgtatgctatatacatata 1770
Db 1441 GAGTACACATATTAGCATGTACCGTTAATTAATCAGCTTCCTTATGTCTATATCTACATACA 1500
OY 1771 taacacaggtatctactataagaataaaggaagaaataaataaagattaaact 1825
Db 1501 TACACAGGTATCTACTATTAAGAAATTAAGAAAGAAAATTAATTAACGATTAAACATT 1555

```

RESULT 3

```

SCYJR116W 1555 bp DNA PLN 11-AUG-1997
LOCUS SCYJR116W 1555 bp DNA PLN 11-AUG-1997
DEFINITION S.cerevisiae chromosome X reading frame ORF YJR116W.
ACCESSION 249616 Y13136
VERSION 249616.1 GI:1015834
KEYWORDS
SOURCE
ORGANISM

```

```

baker's yeast.
Saccharomyces cerevisiae
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; Saccharomycetaceae; Saccharomyces.
REFERENCE
1 (bases 1 to 1555)
AUTHORS Rose,M., Koetter,P. and Entlian,K.D.
JOURNAL Unpublished
TITLE MIPS.
AUTHORS MIPS.
REFERENCE 2 (bases 1 to 1555)
JOURNAL MIPS.

```

```

FEATURES
source
1. 1555
/location/Qualifiers

```

```

1. 1555
/location/Qualifiers
Submitted (25-SEP-1995) Data collected by MIPS on behalf of the
European yeast chromosome X sequencing project. MIPS at the
Max-Planck-Institut fuer Biochemie, Am Klopfersplitz 18a D-82152
Martinsried, FRG; E-mail: Mewes@mips.emblnet.org
Location/Qualifiers

```

CDS

```

1. 1555
/organism="Saccharomyces cerevisiae"
/db_xref="taxon:4932"
/chromosome="X"
374..1213
/note="ORF YJR116W"
/codon_start=1
/protein_id="CA89646.1"
/db_xref="GI:1015835"
/db_xref="SWISS-PROT:P47153"
/translation="MNASTTAAIGLTSPEKLSFFPHSSMLLAHLHEIFSEFYQ
LAFSVAPFLNKVVRKHYTTIRDPDLKIDENVNVAICAVNSNTVLPITPMHY
NVTYTDYSMSVSSLSAGVFTMDLTCVREKLYGLEFGHAISYVYMLLSRPG
OPWIGRELITASPTFVNINMFIMQCNKSKNSIPLVNFVNGLLMTVFVVRICMG
SIASALLEROMKVRDELPRFSAVTMSLNFMLNLVLPWKMRIRAKKLAKPAPTS
KLD"
BASE COUNT 458 a 383 c 259 g 455 t
ORIGIN

```

```

Query Match 18.7%; Score 342; DB 8; Length 1555;
Best Local Similarity 100.0%; Pred. No. 8; 9e-60;
Matches 342; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 1 acctactttttctatcttcaacaagaacgcttacacacacacacatcatct 60
Db 1214 ACCTACCTTTTTTTCTATCTTCAACACGAAAGGCTTACACACACACATCATCT 1273
OY 61 acatacatatatacaataatatacatatgtaaacctgtatctatctatcaacaa 120
Db 1274 ACATACATATACATACAAATATACATATATGTAAACCTGTATATTCCTATTACCAA 1333
OY 121 aagagcaattaaactttccctctttcttaagctatctactcaaaaacttaact 180
Db 1334 AAGAGCAATTAACCTTTCCCTCTTTTCTACGCAATTAACCTCAAAAACCTAATTCCT 1393
OY 181 tgcgtcgtcttgccatctttctccagaaaaaacctgcgggaaataaaaaaaagac 240
Db 1394 TCGTCTGCTTGTGCGCATTTTCTCCAGAAAAAATTCGACGGGAATTAATAAAAAAAGAC 1453
OY 241 aacgacacagagaaaagltcgcgaattataaacactctataatacaagaaaagga 300
Db 1454 AACGAACAAGAGAAAAGTTGCGGAATATTAACCACTTCTATATTAACAGGAAGAAGA 1513
OY 301 aggaataaagaggaatagaaaactgcagccttatct 342
Db 1514 AGGAATAAAGAGAGAAATAGAAAACGACAGGCTTTATTC 1555

```

RESULT 4

```

NCB11N2/c 99360 bp DNA PLN 08-FEB-2001
LOCUS NCB11N2 99360 bp DNA PLN 08-FEB-2001
DEFINITION Neurospora crassa DNA linkage group V BAC contig B11N2.
ACCESSION AL513444
VERSION AL513444.1 GI:12718373
KEYWORDS
SOURCE
ORGANISM

```

```

Neurospora crassa.
Neurospora crassa
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
Sordariales; Sordariaceae; Neurospora.
REFERENCE
1 (bases 1 to 99360)
AUTHORS Schulte,U., Algrn,V., Hohnsels,J., Brandt,P., Fartmann,B.,
Holland,R., Nyakatura,G., Mewes,H.W. and Mannhaupt,G.
JOURNAL Unpublished
TITLE MIPS.
AUTHORS MIPS.
REFERENCE 2 (bases 1 to 99360)
JOURNAL MIPS.

```

```

FEATURES
source
1. 99360
/organism="Neurospora crassa"
/db_xref="taxon:5141"
/chromosome="2"
1..10847
/note="Overlap to BAC B2F7, EMBL:AL451013, please refer to
this entry for analysis and annotation"
complement(8202..11027)
/gene="B11N2.10"
8202..11027
/gene="B11N2.10"
complement(8202..11027)
/gene="B11N2.10"

```

```

Submitted (07-FEB-2001) MIPS, at the Max-Planck-Institut fuer
Biochemie, Am Klopfersplitz 18a, D-82152 Martinsried, FRG; E-mail:
mannhaupt@mips.biochem.mpg.de Project Coordinator: Ulrich Schulte,
Institute of Biochemistry, Heinrich-Heine-University, D-40225
Duesseldorf, E-mail: ulrich-schulte@uni-duesseldorf.de
this contig is an assembly of BAC 11N2 from 1 to: 91660, strain
OR74A, and cosmid 65G7 from 91661 to: 99360, cosmid library
P10R15T6Xh, strain 74-OR-23-1A; BAC and cosmid clones are available
at the Fungal Genetic Stock Center, http://www.fgsc.net
Sequencing was performed by MWG Biotech AG, Ebersberg, Germany,
http://www.mwgda.com
Information on performance of analysis and a more detailed
annotation of this entry and other sequences of linkage groups II
and V can be viewed at:
http://www.mips.biochem.mpg.de/proj/neurospora.
Location/Qualifiers

```

FEATURES

```

misc_feature
1..10847
/note="Overlap to BAC B2F7, EMBL:AL451013, please refer to
this entry for analysis and annotation"
complement(8202..11027)
/gene="B11N2.10"
8202..11027
/gene="B11N2.10"
complement(8202..11027)
/gene="B11N2.10"

```



```
CDS
/number=1
complement(8202..11027)
/gene="B11N2.10"
/note="similarity to predicted rna binding protein, homo sapiens, TREMBL:AB016092.1"
/codon_start=1
/product="conserved hypothetical protein"
/db_xref="GI:12718374"
/translation="MTRSPSSSPNLSHALAVTPRSPPSPSTSPSPRRYRENDLLS
NLTPRTALDITRNPTGEFRACMASASEPTQAFALRAIASNKIOEWLDELSSMPWPG
ASSAGFQVPLKRRKREKSEATOLSKYCSGLPAEVAKYKREYEIKPMKEIDIEI
KNHLYRNHIVPLSAFGSPILGSHRPLPSYSHMDLTFELATIMOTLPNLRSLRN
SMSRLILVLRPAFLDTLADGEIALOSRGVGLDGNARAIOSRPSPSMAEFNVI
KSYLKKVAKAGDLDVMDLSEGWETTLPEMDLRESLEKQGEVNAACEKRIKOA
NLKIDATTSIDTAKLDDERKQOQTNPAPNAVALTVDDQSDDLTDHKTINQV
PLIKVHVPLEGIDENDAEKGSNDSWFVDETAPETNDESLITEERPLGGFPGSDID
GASAGMPPSPNIVASSKGRPHSSLSYATSDGEYELDPDSDPFDNLHDLFG
FPRSRSNFSSTSTIVFGRALOEPTDRSRSSKDLGSLGDSVSPSPSPRS
KSLRSLSRGSDMPVAELPDDESNTLSSAVNENGPESKSNAGOSDVKLDLLOOI
SELESYPARKLTSERPINLNPDPPTALNRKSSKRELNRSSOISLSTSTSSRAG
TPSFTLLAPRSKGRKORNSHRDKITHLTRNNGEAPILFTRCVGENGERVAVRIG
GWMADLGEYLRVESHRRNVALEKVEVMKPRVIAAGSSSPTRPPESTRTORQSP
DTPLNVRKTRKPAEASEENAGTKATPKTPFPVAGKATAVSSPRAPVARSISITR
SRSSRLSHRGEDEIALGLAGPRTKNTIEMSEEDKAWEDKAKVQNASGNRQVSDSP
ATAEGKGDGKIGTKRVRKROGLP"
complement(join(12353..12531,12616..12798,12825..13053))
/gene="B11N2.20"
complement(12353..12531)
/gene="B11N2.20"
/number=1
12353..13053
/gene="B11N2.20"
complement(join(12353..12531,12616..12798,12825..13053))
/gene="B11N2.20"
/product="hypothetical protein"
/codon_start=1
/protein_id="CAC28683.1"
/db_xref="GI:12718375"
/note="MEAGVVGSGSHRPMWSDCLITANMLPDGDLIMRSSVLASGR
SDSSSAHPHTTHPQASLIQEOVGTCAVIVVHSDHVPARSWSAPVHWPVAP
PMAFWSKAPRAKRAHVVDHPKGDGEAGIPEPSMRDGTVPATEPSRORAAEQEPAG
AEDNSQSTQSDRPSDRVOIEMSGPRKRVRIPEG"
complement(12532..12615)
/number=1
12616..12798
complement(12616..12798)
/gene="B11N2.20"
/number=2
12799..12824
complement(12799..12824)
/number=2
12825..13053
complement(12825..13053)
/gene="B11N2.20"
/number=3
complement(join(14115..16258,16397..16536,16654..16958))
/gene="B11N2.30"
14115..16958
/gene="B11N2.30"
complement(join(14115..16258,16397..16536,16654..16958))
/gene="B11N2.30"
/note="strong similarity to CUTINASE TRANSCRIPTION FACTOR
1 BETA, Fusarium solani, SWISSPROT:CHI1B_FUSO"
/codon_start=1
/product="probable CUTINASE TRANSCRIPTION FACTOR 1 BETA"
/protein_id="CAC28684.1"
/db_xref="GI:12718376"
/translation="MSDSGSDQCTSPSPSTENESTGVCSTPCEKEPSSLARIEA
TPSKRPSKALPADSONPRAKRRARACLTICNRKRCVVEGGLCAQCNKKGPP
KNMLAGQAVNGVAAGVEFGARIGIHOPLISSOQLAHANASIMADAOFTOGMLN
NPISPTESARLQNPOLMTMPCNTINPSTASATTPPTPTISVSTPSOPLQONLQPK
PNRTVSPNIDNFRKPLPRKDAVVOYIMAKGATITPPPEYONAMLKATIEYVPIPL
LELRQFLTAHNSGQGISLVLYOAVMFGSNFVAKYITPDEAGLGSRRNARELEMR
TRVLYDDCVKRLDLVALLLMTYWTPEPDOKDTHMIGVAISLGLITIGIRNPAN
LAMPAPQKLLMKRIWCCFMRDLIALGMRRTRIKDEDFVMLERSDEIYELPAD
```



```

gene      /chromosome="1"
           /clone="cosmid c3h1"
           /map="tl"
           complement(join(1..410,455..490))
           /gene="SPAC3H1.01c"
           /note="orf3"
           complement(join(<1..410,455..490))
           /gene="SPAC3H1.01c"
           /note="SPAC3H1.01c, len:148"
           /codon_start=1
           /product="Origin recognition complex-related protein
           subunit 3."
           /protein_id="CA92254.1"
           /db_xref="GI:1103502"
           /db_xref="SWISS-PROT:Q10067"
           /translation="MSA1LODVSQKTYSVKPKPTESLGNESSEGLFVPLIGKEPI
           EYVLRNLCITKVLNLEIEITHTTYSNARVLRQICEHASPEVDGLRTAIVCDK
           SSNSMHLKFEQLDKDMEKIGKMFVNLNKLAKHRLKSCFVIGGS"
           complement(411..427)
           /note="ctactaattgtag, splice branch and acceptor"
           complement(449..454)
           /note="gtaaagt, splice donor sequence"
           complement(914..4024)
           /gene="SPAC3H1.02c"
           complement(914..4024)
           /gene="SPAC3H1.02c"
           /note="SPAC3H1.02c, len:1036, SIMILARITY: Saccharomyces
           cerevisiae, Y0J8_YEAST, hypothetical 118.4 kd protein,
           (1037 aa), fasta scores: opt: 1381, E():0, (29.3% identity
           in 1068 aa)"
           /codon_start=1
           /label="SPAC3H1.02c"
           /product="possible zinc metalloproteinase"
           /protein_id="CA92253.1"
           /db_xref="GI:1103503"
           /db_xref="SWISS-PROT:Q10068"
           /translation="MISDTSFROISERILNTYVNMKINDATGFSVAVKPTSRLO
           GSFVATEAHNDGCPHTLEHLCFMGSKKYPMNGILTRFAGACGDINACDVNTSY
           ELSAEDGFLRLPVPADHLISPLIDEAFTCYVHINNGEGSSGVYSQMOTQSS
           ETDVFCMKRISQPVTSGYTETGHPSELRKLSIKIRYHKEMVPSMCLITVG
           CINESRLSCASQIVKEILANGIITPTTPMPCSTWVDYIIPILKTVFSEED
           TGSVLAAMGNSADATVFAIEILCEFLSEASVPGQFVEIEDPSCFVYVYL
           RVPSIQLFEDSVLEKINGEORALRLAEFOIDNMKMDYLRTRDOVLTLMELF
           PSSLFMKILITLHYVSGREGKDPNLKMLEYNRLLEMEKMDKLKILKMFVNNY
           TVIALPSEMAENKKENAEQLNKRCLLGSGLGAEKLAERKKEKNEQKLSNLS
           SFRISDESIREFYSTARTARSAGQPDNEVQYITDVSESNPFIQPDHDSFVL
           ATYIDTSMIPSLKPYLSVFAKYLEAAPALSGTIPPIYHEVYQLDRDVTSLVNSL
           FDTSSCGAYETRELLSEIKYTRNEYGVYIINRLAKTVWDRMLYSINQDL
           ADIFOKRDAEFLIPSYEDIRLYNDGSLKSLANTSOEKLIRELRDLNDHSSIFDA
           FOKRMNTMLOHQAIRIHVIGDILKLPQISTWNKLLDSECRNEMELSTFEKNTYK
           POERSSSLTVIPMPNNESSDLVFSIPGTTSWDPSLPYVILANTLGLMDGFRNAT
           IRSGGLAYGFMNSIDVGGVLYICINTSPDYKAWASSRDLVKLISGEVAVSSFDLE
           SAKCVSYISVSELMNAIYSSKNSTFLSLKGLAKDDEKHLKLEVKQVTLQDFLEGK
           IYCLPFFSSNNILAVITSSIAKIEQVEQFEFGFNVAVESLHIEIQIESSEDLVY
           GGNDDN"
           complement(3539..3928)
           /gene="SPAC3H1.02c"
           /note="Match to PF00675 Peptidase_M16, Insulinase
           (peptidase family M16) score 21.57"
           4683..5123
           /gene="SPAC3H1.03"
           4683..5123
           /gene="SPAC3H1.03"
           /note="SPAC3H1.03, len:146"
           /codon_start=1
           /label="SPAC3H1.03"
           /product="hypothetical protein"
           /protein_id="CA92256.1"
           /db_xref="GI:1103504"
           /db_xref="SWISS-PROT:Q10069"
           /translation="MSLVAYDSEEDQSTLVNENNDIKGSEEPHAKTIPNSKAEVDT
           ELEKKIKQFIKLKAKGIHFHTRLSENENFRNPKLLDNLQDFLDIKPRGTWISKMDMD

```

```

gene      PTDFHKNVVASALSKSODEMIARRENYOKHDKREIKFQTSQHPK"
           complement(join(5402..5560,5662..7308))
           /gene="SPAC3H1.04c"
           complement(join(3402..5560,5662..7308))
           /gene="SPAC3H1.04c"
           /note="SPAC3H1.04c, len:601, SIMILARITY: Candida albicans,
           094022, hypothetical 69.0 kd protein., (595 aa), fasta
           scores: opt: 1014, E():0, (40.0% identity in 508 aa)"
           /codon_start=1
           /label="SPAC3H1.04c"
           /product="hypothetical protein"
           /protein_id="CA92257.1"
           /db_xref="GI:1103505"
           /db_xref="SWISS-PROT:Q10070"
           /translation="MLESRLAKIARIARQVARKPYLISYHVRASPNALLKQTVLOS
           EKSEPTILPRLAARNTSNGLSRTTPIITIKQISVRSYSSGDENPEITLKNESNOS
           GVRKAMPFVRLKMKSLFLFQNKPLVDVNTAFESVWLVAHIWVGTTFESLLY
           TLNTVSAQELLGRWIGQMLTKNTGFQFESALVPNNRKGLIFENKLSVLRPDTLNG
           IGADPNPNKSDPEKEKYMALRRKYDSNEPDTALSQCNTOFELSIDKADVFESFAR
           LNKGIYKELQKRGVVDRRFIEMDPSDPDRYRKRHHMGDEIEKFLIEDRLVTL
           LQPKGFEPVSVFPCFLPRLRKQMLFYDLMNAKMTINDGSDDELVTISPSSEPSNF
           LKVGKQIDDMRHSRLRDNAIDHLNNGVSGARNTINDGSDDELVTISPSSEPSNF
           OKAVISLMDKLKKEKDEVDYKDVHVNQALHNPRAVPIFTNQVSYINNALIRPIT
           AYINSTRFTIPLICVSKPLSDFDGSMTFYDSGVLOEISAQVYESFARDVLNDEIRK
           RIQVGYWSLRFLHLVLSLOELAPTTSSISNE"
           complement(5561..5576)
           /note="ctactttagtag, splice branch and acceptor"
           complement(55656..5661)
           /note="gtaaagt, splice donor sequence"
           join(8226..8330,8401..9720)
           /gene="SPAC3H1.05"
           join(8226..8330,8401..9720)
           /gene="SPAC3H1.05"
           /note="SPAC3H1.05, len:474, SIMILARITY: Saccharomyces
           cerevisiae, ST24_YEAST, caax prenyl protease 1, (453 aa),
           fasta scores: opt: 896, E():0, (40.5% identity in 454 aa)"
           /codon_start=1
           /label="SPAC3H1.05"
           /product="putative caax prenyl protease 1"
           /protein_id="CA92258.1"
           /db_xref="GI:1103506"
           /db_xref="SWISS-PROT:Q10071"
           /translation="MSPGLCKEISYIQATPKPTTSFANCCMGILQHLMLHIDIP
           GFPMKVIAGSISKVMADYLRRROYVLYLRKSPALIAHEVKEKYOKALSVAROK
           SMESTIYSTPLADLLILIKYDGLSYLMTNTRKPPMMKIDLAASSRSRLTSSICVF
           MGLTLESRLQIQPFLNLTVEIEKQSTIKTYIDILKLSLGGLSMYYQVEA
           FVKLTLEFGDNFIYAWGAYIVFELIQTAPSLIMPLFYKFPYLENGSRLTQVEEA
           ASINPFLKLVIVIDASRSRSHNAFFGLPWNKGVLEFDLVKNHTPELIALIGHEL
           GHVMSHNLINTIIDYDMSLFLHLFAFIRNNLSYFNFTERKPIYVIGLLFSDEL
           GPLSSITLTPASNKYSRLCEYQADAFQAGVAKDLGGLIRIHDNLSLPFEDSLYTS
           YVSHPTLVORLNAIDVTTLKNN"
           8331..8336
           /note="gatal, splice donor sequence"
           8386..8400
           /note="ctaacgttttag, splice branch and acceptor"
           8647..8691
           /gene="SPAC3H1.05"
           /note="PS00678 Beta-transducin family Trp-asp repeats
           signature"
           9106..9690
           /gene="SPAC3H1.05"
           /note="Match to PF01435 Peptidase_M48, peptidase family
           M48 score 388.94"
           9280..9309
           /gene="SPAC3H1.05"
           /note="PS00142 Neutral zinc metalloproteinases,
           zinc-binding region signature"
           complement(10343..12112)
           /gene="SPAC3H1.06c"
           complement(10343..12112)
           /gene="SPAC3H1.06c"
           /note="SPAC3H1.06c, len:589,
           SIMILARITY: Schizosaccharomyces pombe, 042922, hypothetical

```


| | | | | |
|------------|---|-------------------------|-------|-------------|
| Oy | 1150 | ccatcgaactccaagagaatt | 1169 | |
| | | | | |
| Db | 21 | CCGTTTTTCTCTAAACGGAT | 2 | |
| RESULT | 8 | | | |
| LOCUS | AR155763/c | | | |
| DEFINITION | AR155763 | 686 bp | DNA | PAT |
| ACCESSION | Sequence 929 | from patent US 6239264. | | 08-AUG-2001 |
| VERSION | AR155763 | | | |
| KEYWORDS | AR155763.1 | GI:1512816 | | |
| SOURCE | Unknown. | | | |
| ORGANISM | Unknown. | | | |
| REFERENCE | Unclassified. | | | |
| AUTHORS | 1 (bases 1 to 686) | | | |
| | Phillippsen, P., Pohlmann, R., Steiner-Lange, S., Mohr, C., Wendland, J., | | | |
| | Knechtle, P. and Reibischung, C. | | | |
| TITLE | Genomic DNA sequences of ashbya gossypii and uses thereof | | | |
| JOURNAL | Patent: US 6239264-A 929 29-MAY-2001; | | | |
| FEATURES | Location/Qualifiers | | | |
| source | 1..686 | | | |
| | /organism="unknown" | | | |
| BASE COUNT | 179 a | 173 c | 169 g | 165 t |
| ORIGIN | | | | |

| | Query Match | 7.5% | Score 136 | DB 6 | Length 666 |
|----|-----------------------|---|-------------------|----------|------------|
| | Best Local Similarity | 52.1% | Pred. No. 7.8e-18 | | |
| | Matches 354 | Conservative 0 | Mismatches 320 | Indels 6 | Gaps 2 |
| QY | 430 | ctggaagaacgaatgatgatgaatcatttccataaatacgaagaactactcccgcccaag | 549 | | |
| Db | 675 | CTGGAGGGCGTATTGATTAAGAAACCAATGCAAAAGACGACGCGTAGAGGGGCCAAG | 616 | | |
| QY | 550 | gccaaagtctccatllttcgtgtagcgtctataacctagcccaagaagctagtttccataa | 609 | | |
| Db | 615 | TAAACGTCACCGTATGTTAGAGATGTGGTGTTCCTCCGGGACTAAATCTGTATATGATTAA | 556 | | |
| QY | 610 | taagaccccttcocccaatctgacacatgacggtttcttcttaatgaaagcagtcctgca | 669 | | |
| Db | 555 | TACAGACTGGCTGCCACGAGATGTGGAACTAGGAGTGGCGGTAGGTACGCG---CATGCC | 499 | | |
| QY | 670 | gtcagaattcatalatggtctcacctgctgcacagaagttatgcttcttgcctctatcc | 729 | | |
| Db | 498 | GCCATGCTGTGGCCAGTATCGACCATCTCGACGTGGCTGTACTTTTGTATGCTTACCTA | 439 | | |
| QY | 730 | agttgcttacctggttgatlltgcgaactctctactataagccatlltgcctggaaga | 789 | | |
| Db | 438 | CAGCTGAACCTGGTGGCAGGGGTTGTTGGAGACTATTAATCAACATTTGGTCTCAGGAA | 379 | | |
| QY | 790 | aaatttgatttccaataaattgacccgtcccaactatgatacagatatatcaagaagctg | 849 | | |
| Db | 378 | AAATTCCGGGTTCAACAACTCCAGCGTCAAGCATGTGGCTAACGGACACAGCTAAAGTGT | 319 | | |
| QY | 850 | actltgacgtatgcctatgtgtgcgccaactcccttaacctgttcccttaagaatcttgataa | 909 | | |
| Db | 318 | ATGATCTCTCTCATGATATACCACTCTCCAGCCCTATCGCGTGTGAAGTAAATTGAAGA | 259 | | |
| QY | 910 | ttccctactgatttcccttctgttatactatgcttctctgttctgttgcctccaatcttagc | 969 | | |
| Db | 258 | TTTCGTGACTGGCTTGTGTAGCTAGCTCTATACCTGATGCTGTTCTTATCTTGGTGTG | 199 | | |
| QY | 970 | atgaacaacatcttcagcttcatactacgcacatgtttaaataagttacacccaattgaagac | 1029 | | |
| Db | 198 | ACTGCTCTTCAGCCAGTT---TTACAGCTTTATTTCACCAAGCTACCCCGTTGGAAAC | 142 | | |
| QY | 1030 | ggtagaactgaaaaaatcctatgtgaagtttgcgcgaatagatgttggttccctctagataag | 1089 | | |
| Db | 141 | GGGAGAGCTGAAGACGTGCTGATTTGTGAAATTAAGCAAAACGATCACTTCCCATGTGACAA | 82 | | |
| QY | 1090 | attttgcatctgacggtctcaaaaagatcttctcaatcaagaacgatatttcaaggttg | 1149 | | |

| | | | |
|------------|--|---|-----------------|
| Db | 81 | ATCTACCTTTTCGGACGGCTCGCGCCGCTCTGCGCAGCTAAATGCGTATTTCACGGGGCTA | 22 |
| QY | 1150 | ccattccacctccaagaagaat | 1169 |
| | | | |
| Db | 21 | CCGTTTTCCTCTAAACGAT | 2 |
| RESULT | 9 | | |
| E66288/c | | | |
| LOCUS | E66288 | 686 bp | DNA |
| DEFINITION | Genome DNA sequence of Ashbya gossypii and use thereof. | | PAT 07-FEB-2001 |
| ACCESSION | E66288 | | |
| VERSION | E66288.1 | GI:13020629 | |
| KEYWORDS | JP 1999225770-A/928. | | |
| SOURCE | unidentified. | | |
| ORGANISM | unidentified. | | |
| REFERENCE | 1 (bases 1 to 686) | | |
| AUTHORS | Peter,P.R.P., Steiner,C.M., Juergen,W.P.K.K. and Reblshun. | | |
| TITLE | Genome DNA sequence of Ashbya gossypii and use thereof | | |
| JOURNAL | Patent: JP 1999225770-A 928 24-AUG-1999; | | |
| COMMENT | NOVARTIS AG | | |
| | OS PAG1577RP | | |

| Query Match | Best Local Similarity | Matches | Conservative | Mismatches | Indels | Gaps |
|-------------|---|---------|--------------|------------|--------|------|
| QY 490 | cttgaagaagcaaatgatatgataagcaatttccataaatcaagaactactcccgagccaag | 549 | | | | |
| Db 675 | CTGGAGGGCGTATTGATTAAGAAAGAAACATGCAGAAAGACGAGGGGTACGAGCGGCCACAG | 616 | | | | |
| QY 550 | gcacaagttctccatttccgtgtgacgctctataacctagaccacaagaagctagtttccatcaa | 609 | | | | |
| Db 615 | TAAAGCTAACCCGTATGAGTAGAGATCTGGTCTCTGGGACATAAATCTGTGATGATTMAAG | 556 | | | | |
| QY 610 | tacgaccccttcccttaaatctgcaacgtgcgtttctttatgaaagcagctcctgcga | 669 | | | | |
| Db 555 | TACGACCTGGCGCCACGCGATGTGAACCTAGAGAGTGGCGGTAGGTCACGC--CATGCC | 499 | | | | |
| QY 670 | gtcagatttcatcagctccacgtccgacagagttatgctctcttggctcttacc | 729 | | | | |
| Db 498 | GCCATTGCTGGTGCAGATATGCACCATCTCGAGTCCGCTGACTTTTGTGATGCTTACTTA | 439 | | | | |
| QY 730 | agttgtctacactgtgttattgttgcacactctctactactaagcaatttgcctcggaa | 789 | | | | |
| Db 438 | CAGCTGAACCTGGTGGACGGGGTGTGGAGAGTATTACATCAACTTTGCTCGAGGAA | 379 | | | | |
| QY 790 | aaatttggttccaataaatgacccgtccaactatgatacgataatgataagagctg | 849 | | | | |
| Db 378 | AAATTGGGTTCAACAAGTCCACGCTCAAGCTATGAGCTAAGGACCAAGCTAAAGTGTTT | 319 | | | | |

| QY | 850 | acttggcgatcgtcgtggtggtcccaatcccttaccgcttccttaagaatcttgataa | 909 |
|------------|---|---|-------------|
| Db | 318 | ATGATCTCCTCCATGATTAATACCACTCTCTGACGCTTAAGCGCTCTTAAGTAAATTAAGAAG | 259 |
| QY | 910 | ttccctactgaatttcctttgtgtacattatggtctctctgttgcgtgtgtccaaatcttaacc | 969 |
| Db | 258 | TTCCTGACTGCTTGCTGAGCTACGCTCTCTATNCTGATNGCTGTCTTCTCATCTGGTGTG | 199 |
| QY | 970 | atgacaatcatctccagcttcattcatcgtcccatggtttaataagttactactcatcttgagac | 1029 |
| Db | 198 | ACTGCTTCTTCAGCCCGT---TTCACACTTTATTTCACACAGCTCACCCCGTTGGAAAGAC | 142 |
| QY | 1030 | gttgacacgaataaactcatttgaagtttgccgagtaagttggttccctccttagaataag | 1089 |
| Db | 141 | GGGGGCTGAAAGACCTCATTTGTGCAAAATGACAAACAGCATCACTCCCATGGACAA | 82 |
| QY | 1090 | attttgcatcgtacgctgcataaagaatcttcattcatcaacgcatatttcacagtttg | 1149 |
| Db | 81 | ATCTACCTTTGGAGAGGCTCGCGCGCTGTGGGCACTCAANTGCTATTTCACGGGGCTA | 22 |
| QY | 1150 | ccattaccctccaaagagaat | 1169 |
| Db | 21 | CCGTTTCTCTCTAAACGGAT | 2 |
| RESULT | 10 | | |
| LOCUS | SCYR118C | | |
| DEFINITION | SCYR118C | 935 bp | DNA |
| ACCESSION | S.cerevisiae chromosome X reading frame ORF YR118C. | | 11-AUG-1997 |
| VERSION | 249618.1 | GI:1015838 | |
| KEYWORDS | | | |
| SOURCE | | | |
| ORGANISM | | | |
| REFERENCE | | | |
| AUTHORS | | | |
| JOURNAL | | | |
| REFERENCE | | | |
| AUTHORS | | | |
| JOURNAL | | | |
| TITLE | | | |
| FEATURES | | | |
| source | | | |
| CDS | | | |
| BASE COUNT | | | |
| ORIGIN | | | |

| | | | | |
|-----------------------|----------------|---|----------|------------|
| Query Match | 6.6% | Score 121 | DB 8 | Length 935 |
| Best Local Similarity | 100.0% | Prod. No. | 8.7e-15 | |
| Matches 121 | Conservative 0 | Mismatches 0 | Indels 0 | Gaps 0 |
| Oy | 1705 | tctataagtagaacattatagcatgtaacggttaaatccagcttcggttatgcttaaccta | 1764 | |
| Db | 1 | tctatagagtcacacattatagcatctaccggttaaatcagcttcggttatgcttaaccta | 60 | |

| | | | | |
|----|--|------|--|------|
| QY | | 1765 | caccacaaacaggtatctactctaagaataaaggaaaqaaaaaaaaaataaacgatttaaacat | 1822 |
| | | | | |
| Dδ | | 61 | CATCATACACAGGTATCTCTTAAGAATAAAGAAAGAAAAATAATTAACGATTAAACAT | 120 |
| QY | | 1825 | t | 1825 |
| | | | | |
| Dδ | | 121 | T | 121 |

| FEATURES | REFERENCE | LOCUS | DEFINITION | ACCESSION | VERSION | KEYWORDS | SOURCE | ORGANISM | TITLE | JOURNAL | MEDLINE | REFERENCE | AUTHORS | TITLE | JOURNAL |
|----------|-----------|---------|-------------|-----------|------------|----------|--------------|---|---|---------------------------------------|----------|---------------------|--|-------------------------|---|
| AF064867 | 1959 bp | 1859 bp | 28-OCT-1998 | AF064867 | GI:3800768 | human. | Homo sapiens | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. | Dual roles for Ste24p in yeast a-factor maturation: NH2-terminal proteolysis and COOH-terminal CAX processing | J. Cell Biol. 142 (3), 635-649 (1998) | 96365461 | 2 (bases 1 to 1859) | Tam, A., Nouvet, F.J., Fujimura-Kameda, K., Siunt, H., Sisodia, S.S. and Michaelis, S. | Submitted (12-MAY-1998) | Cell Biology and Anatomy, Johns Hopkins University School of Medicine, 725 N. Wolfe St., Baltimore, MD 21205, USA |

| FEATURES | SOURCE | location/Qualifiers |
|---------------------------|--------|--|
| gene | CDS | 1..1859 |
| | | /organism="Homo sapiens" |
| | | /db_xref="taxon:9606" |
| | | /tissue_type="B cells; fetal brain" |
| | | 1..1859 |
| | | /gene="STE24" |
| | | 124..1551 |
| | | /gene="STE24" |
| | | /note="similar to Saccharomyces cerevisiae Ste24p" |
| | | /codon_start=1 |
| | | /product="CAAX prenyl protease" |
| | | /protein_id="AAC68866.1" |
| | | /db_xref="GI:3800769" |
| | | /translation="MGMMASLDALMWPAPKRIFGAVLLFSMTVYIMETFLAQRORRI YKTTTHVPELPGQIMDESETEFSRLYOLDSTFSFMGLSSEFGTLILFLGGIPRYLM RSGRPGVAGFGEYEITOSIVFLATLTFESALTGPMSLVYNFVEIEKHGFMQDTL GAPMDLKKRVVTCOILLPVSSLLYTIKIGDYEPTIAMLFTLVLYLVLYTYADY IAPLDFKFTPLPEGKLEKEIEVMKKSIDFELTKYVYVEGSRSSHSAHYIGFFKNRR IVLEDTLLEEYSLVANKDIOEDSGEPNNEEGNSEELIKAVYKNNKQCKNEEVLAVLG HEHGPMKLGHTVKNIIILSQNNSEFLCFPLFAVLIGRKELFAAFGFDYDQPTLIGILLIF QIFESPNEVIESFLTVLSRRFEPOADAFAKKLGAKDLYSALIKLNKNDLGFVSDOM LFMSMHSHPPLEERLOALKTMKOH" |
| BASE COUNT | 526 a | 336 c 405 g 592 t |
| ORIGIN | | |
| Query Match | 6.4% | Score 117.4; DB 9; Length 1859; |
| Best Local Similarity | 50.5% | Pred. No. 4.6e-14; |
| Matches 370; Conservative | 0; | Mismatches 351; Indels 12; Gaps 3; |

[illegible]

Qy 591 aaagatgtttcatcaataacagacccctccctaaatctgycacatgagccgttctt 650
 Db 387 AGGCACTCTTATTCTTCTTCTTGAGGAATACCTTATCTCGAGACATTTTCGACGGTT 446
 Qy 651 attgaatcagtcctgcagtcagtcagatctcacaatgctccactgycagagattatg 710
 Db 447 CT--CTGGTTATGCTGGCTTTGGACCAAGATATGAGATCACTCACTGCCCTGGTTCTG 504
 Qy 711 ctctctggctctctatccagttgtctacactggttgatctgcacccctctactatag 770
 Db 505 CTGTTGGCTACACT-----TTTCAGTGCATTGACTGCTTGGCCATGAGCTTTTATTA 557
 Qy 771 ccaattgtccctgagaagaaattgtgtctcacaataatcagccgcacactatgatac 830
 Db 558 TACTTTTGTGATAGAGAAAGAAACATGGCTTCATACACAGACTTTGGGGTTCTTATGAA 617
 Qy 831 cgaatgatcaagagctcgaacttggtggtatgctatggtggcccaatcccttaccgtt 890
 Db 618 AGATCCAAATCAAGAAATTTGTTGCTCAGTGTATTTTGGCTGCTGCTTCACTTCT 677
 Qy 891 ccttaagatccttgataaattccctacagatctcctctgtgtacatataatgctcttct 950
 Db 678 ACTTTACATTATTAATAATGGGGGTGACTATTTTATTTATGCTGGCTGTTTACATT 737
 Qy 951 cgtgtccaatcttagcacaatcacaatccagctctcacaatcagccatgattcaataa 1010
 Db 728 AGTTGTGCTCTGCTGTTCTTGACAAATCTATGCTGATATATTTGGCCCTTATTTGACAA 797
 Qy 1011 gttcactccatgtgagcaggtgaaactgaaataatctatgaaagtgtggccagatag 1070
 Db 798 ATTTCACCTCTGCTGAGGGGAAACCTTAAGAAATTTGAAGAAATTTGCAAAAGATAT 857
 Qy 1071 tgggttccctcctagaatattgtgtcatgtgagcgtccaaagaattctcattcaaa 1130
 Db 858 TGACTTCTCTTTGACGAAGGTGTATGTGTGGAAGATCTAAACCTTCTCCACAGCAA 917
 Qy 1131 cgcataatcacaagttgtccatccctccaagagaattgtgtgtccacactatg 1190
 Db 918 TGGTATTTTATGCTCT---CTTCAAGAACACAGGATAGTTTGTGACACTCTACT 974
 Qy 1191 gaacagtaattct 1203
 Db 975 AGAAGACTACTCT 987
 RESULT 12
 AB016068 2963 bp mRNA PRI 21-JAN-2000
 DEFINITION Homo sapiens mRNA for Ste24p, complete cds.
 AB016068
 VERSION AB016068.1 GI:3721863
 KEYWORDS Ste24p.
 SOURCE Homo sapiens brain tissue_11b:fetal brain cDNA library cDNA to mRNA.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (sites)
 AUTHORS Kumagai,H., Kawamura,Y., Yanagisawa,K. and Komano,H.
 TITLE Identification of a human cDNA encoding a novel protein structurally related to the yeast membrane-associated metalloprotease, Ste24p
 JOURNAL Blochim. Biophys. Acta 1426 (3), 468-474 (1999)
 MEDLINE 99177429
 REFERENCE 2 (bases 1 to 2963)
 AUTHORS Komano,H. and Kumagai,H.
 TITLE Direct Submission
 JOURNAL Submitted (07-JUL-1998) to the DDBJ/EMBL/Genbank databases. Hiroto Komano, National Institute for Longevity Sciences, Department of Dementia Research, 36-3 Gengo, Morioka-cho, Odu, Aichi 474-8522, Japan (E-mail:hkomano@nlls.go.jp, Tel:81-562-46-2311, Fax:81-562-44-6594)

FEATURES
 source
 Location/Qualifiers
 1..2963
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /tissue_lib="fetal brain cDNA library"
 /tissue_type="brain"
 179..1606
 /gene="Ste24"
 179..1606
 /gene="Ste24"
 /product="Ste24p"
 /protein_id="BAA33727.1"
 /db_xref="GI:3721864"
 /translation="MGWMAISDAIMEMPAKRIFGNVLLFSYTYLMEFLAQRORI
 KYTTHVPELGOIMDSERFKRLVOLDSTSFWSGLTSETGLTLLFGGLPYLM
 RLSEFCGVAGFPEYETITQSLVFLATLFSALTGLPMSLYNFTVEKHGNOQL
 GFEMKDAIKRNVVQCLILPVSLILYIRIGDYFPIVAMLEPLVSLVLTLYADY
 IAPLEDFPTLPBGLKEELVMAKSIDPLTKYVVEGSRSSNAFYGFENKR
 ILEDTLLEESVYNKDIODSGEPNEEGNEEELIKAVKKNKCKGCKNEVLAVG
 HELGMRKLGHTVKNILITISOMNSPLCEFLPAVLGRKELPAFYDQPLIGLITF
 QFISPINEVLSEFLTVLSRFEFOADAFKKAKDLISALIKLNKNDGFEVSOM
 LFSWMHSHPELLERLQALTKMKNH"
 BASE COUNT 847 a 510 c 616 g 990 t
 ORIGIN
 Query Match 6.4% Score 117.4; DB 9; Length 2963;
 Best Local Similarity 50.5% Pred. No. 4.5e-14;
 Matches 370; Conservative 0; Mismatches 351; Indels 12; Gaps 3;
 Qy 471 aacaagttgcccactgctgctgagcagaagaatgaibgaacttccataatcaag 530
 Db 322 AACATCATACACCGCGGCTTAGACAGATCATGAGATTTGAAATTTGAAATTCG 381
 Qy 531 gaactactccgggccaagccaagttccatcttcggtgagctctataactgaagcca 590
 Db 382 ACTGTATCACTGATGAATAAACACATTTAGCTTGTGCTGACAGACTCTATTCAGAGACTGA 441
 Qy 591 aaagatgtttcatcaataacagacccctccctaaatctgycacatgagccgttctt 650
 Db 442 AGGCACTCTTATTCTTCTTCTTGGAGGAATACCTTATCTTGAGACTTTCGACGGCTT 501
 Qy 651 attgaatcagtcctgcagtcagatctcacaatgctccactgycagagattatg 710
 Db 502 CT--CTGGTTATGCTGGCTTTGGACCAAGATATGATCACTACACTCCCTGGTGTCTG 559
 Qy 711 ctctctggctctctatccagttgtctacactggttgatctgcacccctctactatag 770
 Db 560 CTGTTGGCTACACT-----TTTCAGTGCATTGACTGCTTGGCCATGAGCTTTTATTA 612
 Qy 771 ccaattgtccctgagaagaaattgtgttcaataaattgacacgtcccaactatgatac 830
 Db 613 TACTTTTGTGATAGAGAAAGAAACATGGCTTAAATCAACAGACTTTGGGGTCTTCTTCA 672
 Qy 831 cgaatgatcaagagctcgaacttggtggtatgctatggtggcccaatcccttaccgtt 890
 Db 673 AGATGCCAATCAAGAAATTTGTTGTGACTCACTGATATTTTGGCTGTGTTCACTTCT 732
 Qy 891 ccttaagatccttgataaattccctacagatctcctctgtgtacataatgctcttct 950
 Db 733 ACTTTACATTATTAATAATTTGGGGGTGACTATTTTATTTATGCTGCTGTTCACTT 792
 Qy 951 cgtgtccaatcttagcacaatcacaatccagctctcactatgcatgccaatgattataa 1010
 Db 793 AGTTGTGCTCTGCTTCTTGACAAATCTATGATGATATATTTGCCCTTTATTTGACAA 852
 Qy 1011 gttcactccatgtgagcaggtgaaactgaaataatctatgaaagtgtggccagatag 1070
 Db 853 ATTTCACCTCTGCTGAGGAAAGCTTAAGAAATTTGAAGATATGGAAGAGATAT 912
 Qy 1071 tgggttccctcctagaatattgtgtcatgtgagcgtccaaagaattctcattcaaa 1130

Db 913 TGACTTTCCTTGACGAGGTGATGTGTGAGAGAGTCTAAACGCTTCCACAGCAA 972

Qy 1131 cgcataattccagagtttgcattccaccagaagaattgttggttgcacattagt 1190

Db 973 TCCATTATTTTATGCGCTT---CTTCAAGAACAGCAAGATGTTGTTGACACTCTACT 1029

Qy 1191 gaacagtaattct 1203

Db 1030 AGAAGAGTACTCT 1042.

RESULT 13

HSY13834 2966 bp mRNA PRI 01-JUL-1999

LOCUS HSY13834

DEFINITION Homo sapiens mRNA for farnesylated-proteins converting enzyme 1.

ACCESSION Y13834

VERSION Y13834.1 GI:5327058

KEYWORDS facel gene; farnesylated-proteins converting enzyme; gluzincin; metalloproteinase.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 2966)

AUTHORS Freije,J.M., Blay,P., Pendas,A.M., Cadianos,J., Crespo,P. and Lpez-otin,C.

TITLE Identification and chromosomal location of two human genes encoding enzymes potentially involved in proteolytic maturation of farnesylated proteins

JOURNAL Genomics 58 (3), 270-280 (1999)

MEDLINE 99303558

REFERENCE 2 (bases 1 to 2966)

AUTHORS Lopez-otin,C.

TITLE Direct Submission

JOURNAL Submitted (13-JUN-1997) C. Lopez-otin, Universidad de Oviedo, Dept. de Bioquímica y Biología Molecular, Facultad de Medicina, C/ Julian Claveria s/n, 33006-Oviedo, SPAIN

FEATURES

Source

1..2966

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone_lib="LAMBDA DR2"

/clone="A-1.3a"

/tissue_type="Ovary"

34..1461

/gene="facel"

34..1461

/gene="facel"

/codon_start=1

/product="farnesylated-proteins converting enzyme 1"

/protein_id="CAB46277.1"

/db_xref="GI:5327058"

/translation="MGMASLIDALMEMPAERKIFGAVLLFSWTYVLMWTFPLAQRRI YKTTTHVPELGOIMDSSETEPKSRLYQDKSTFSWSGLYSEETGLILLGGIPLYM RLSGRCYAGPEYETIOSLIVFLATLPSALGFLMSLYNTFVIEKGSFNQOOL GEFMKDAIKKEVVOCCILTPVSLITIKIGGDYFFIYAMLFITVIEVLTIVTADY IAPLEKTPPLPEGLKAEIEVMASIDPPLTKRYVVGSKRSSSNATYGFPRNKR IVDPTLLEETSVLNKDIOEDSGMEPRNEEGNSEIIRAKVKNKQCKNEVLAVIG HELGHWKLGHTVKNIIISQMSFLCFPLFAVLIGRKELEFAAGFYDSOPTLGLIIF OFIPSPYNEVLSFCITVLSRREFQADAFARKIGRAKDLVSALLIKLNKDNLGFPVSDM LFSMWHYHPLEETLOALKTMKOH"

BASE COUNT 844 a 515 c 578 g 1029 t

ORIGIN

Query Match 6.4%; Score 117.4; DB 9; Length 2966;

Best Local Similarity 50.5%; Pred. No. 4.5e-14;

Matches 370; Conservative 0; Mismatches 351; Indels 12; Gaps 3;

Qy 471 aacaaagtgcacactgtgcgtgaagacgaattgatgataaactttcataaatcaag 530

Db 177 AACCTATGACACCGAGACTAGACAGATCATGATCTCTGAACATTTGAGAAATCTCG 236

Qy 531 gaactactccggcgccaaagccaagttccacttccatttccgttgacgtatataactgccca 590

Db 237 ACTCTATCAACGTGATTAAGCACTTTCAGCTTGTGTAAGACTATTTACAGACTGA 296

Qy 591 aaagctagtttccatcaataacagaccttccctaaacatccgcaatgacggttctt 650

Db 297 AGGCACTCTTATTTCTCTCTTTGGAGGATACCTTATCTCTGAGACTTTTGGAGGTT 356

Qy 651 attgaatgcagtcctgcagcagatccatataatgctccactgcagtcgacagaattatg 710

Db 357 CT--GTGGTTATGCTGGCTTTTGACCCAGATATGAGATCATGACAGTCCCTGGGTTTCG 414

Qy 711 cttcttgagctcttataccagatttgcctacacctgtgtgttggctccctcttactatg 770

Db 415 CTGTGGCTTACACT-----TTTCAGTGCATTGACTGCTTGGCCATGAGACTTTTATA 467

Qy 771 ccaatttgcctgcgaagaaatttggcttcaataaattgcagcgttccaaactatgatac 830

Db 468 TACTTTTGATGATGAGAAAGAAACATGCTTCATCAACAGACTTTGGGTTCTTCATGAA 527

Qy 831 cgaatagatcaagagtcgacttggcgatgctatctggtggtgccaatcccttaacctgt 890

Db 528 AGATGCAATCAAGAAATTTGTTGACTCAGTGTATTTGTTGCCCTGTCTTCACTCT 587

Qy 891 ccttaagatccttgataaattccctactgatttcccttggctgataatgcttctgt 950

Db 588 ACTTTACATTATTAATAATTGGGGGTGACTATTTTATTTATGCTGCTGTTCACAT 647

Qy 951 cgttgcacaacttgcacatgacatcaatccagcttccatccatgacagtttaataa 1010

Db 648 AGTTGTCGTCGCTTCTTGTACATCATGCTGATTTATTTGCCCTTTATTTACAA 707

Qy 1011 gttcaactcattggagagcgtggaactgaaataatcattgaagtgttgcgataagat 1070

Db 708 ATTCAACCTCTGCTGAGGAGGAAGCTTAAGAAAGAAATGGAAGTATGCAAGAGTAT 767

Qy 1071 tgggtccctcctagataaatttggcttgcatacagcgtcgaagaaatccttccatcaa 1130

Db 768 TGACTTTCCTTGACGAGGTGATGTGTGTGAGAGATCTAAACGCTTCCACAGCAA 827

Qy 1131 cgcataattccagagtttgcattccaccagaagaattgttggttgcacattagt 1190

Db 828 TCGTATTTTATGCGCTT---CTTCAAGAACAGCAAGATGTTGTTGACACTCTACT 884

Qy 1191 gaacagtaattct 1203

Db 885 AGAAGAGTACTCT 897

RESULT 14

E32056

LOCUS E32056 2968 bp DNA PAT 07-FEB-2001

DEFINITION Human AFCL.

ACCESSION E32056

VERSION E32056.1 GI:13026656

KEYWORDS JP 1999113579-A/1.

SOURCE unclassified.

ORGANISM unclassified.

REFERENCE 1 (bases 1 to 2968)

AUTHORS Kristlin,K.K.C.C. and Sazan,A.M.N.N.

TITLE Human AFCL

JOURNAL Human AFCL

COMMENT Patent: JP 1999113579-A 1 27-APR-1999;

SMITHKLINE BEECHAM CORP,SMITHKLINE BEECHAM CORP PUBLIC LTD CO

OS Unidentified

PN JP 1999113579-A/1

PD 27-APR-1999

PF 24-JUN-1997 GB 97304440:7.12-FEB-1998 US 09/022699 PI

KRISTIN K KIRKLY,CHRISTOPHER D SAZAN,ANN M NABU PC

C12N15/09,A61K38/00,A61K39/395,A61K48/00,C07K14/47, PC

C07K16/18,

PC C12N9/64,C12Q1/68,G01N33/15,G01N33/53,C12N15/00,A61K37/02 CC

Db 816 TTGTGCTTACGACACTTTGATTAACCAAGTCGACACTCCGGCATTTGTCGCGCTTAG 875
QY 1229 cccatgaaatcgcgtcactgcaaaaaaacacacgcgttaataatgtaatcatttagtcaat 1288
Db 876 CCCACGAACTTGGACATTTCAAAATGAGCCACAGTACAAAGAACCTGGTCTCCTCAGCAGC 935
QY 1289 tgcacaccttccatcttcccttcccttccacagacatctacagaatatacatcatttaca 1348
Db 936 TCTACATGCTCGTCTTTTATTCCTTTTCAGCCAGTCGCTTTTAACTGATTTATATACC 995
QY 1349 acaccttcggccttcttcttagag 1371
Db 996 GGAGTTTCGGATTTTCCACCTGAG 1018

Search completed: February 12, 2002, 13:20:11
Job time: 9719 sec

THIS PAGE BLANK (25 PTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 12, 2002, 09:32:24 ; Search time 17.02 Seconds

(without alignments)
678.579 Million cell updates/sec

Title: US-09-165-460A-4

Perfect score: 1651

Sequence: 1 MAFSTFLVLYISIVLP.....ISLKDPLQTLVGTGPRRTTL 315

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database: SWISSPROT_39.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|--------------------|
| 1 | 1651 | 100.0 | 315 | 1 | 003530 saccharomyc |
| 2 | 249 | 15.1 | 290 | 1 | PAC2_DROME |
| 3 | 230.5 | 14.0 | 329 | 1 | PAC2_MOUSE |
| 4 | 229.5 | 13.9 | 329 | 1 | PAC2_HUMAN |
| 5 | 202.5 | 12.3 | 271 | 1 | RCE1_SCHPO |
| 6 | 100.5 | 6.1 | 321 | 1 | NU4M_RHISA |
| 7 | 99 | 6.0 | 321 | 1 | MRAY_STRAU |
| 8 | 96.5 | 5.8 | 226 | 1 | ATP6_ANOGA |
| 9 | 96.5 | 5.8 | 534 | 1 | Y255_MYCPN |
| 10 | 96 | 5.8 | 331 | 1 | GPRT_HUMAN |
| 11 | 96 | 5.8 | 426 | 1 | YAS3_SCHPO |
| 12 | 95.5 | 5.7 | 224 | 1 | ATP6_DROSI |
| 13 | 94.5 | 5.7 | 224 | 1 | ATP6_DROYA |
| 14 | 94.5 | 5.7 | 312 | 1 | OLP6_CHICK |
| 15 | 94 | 5.7 | 226 | 1 | ATP6_ANOGU |
| 16 | 93 | 5.6 | 2376 | 1 | YIM6_YEAST |
| 17 | 92 | 5.6 | 715 | 1 | NU5W_NECR |
| 18 | 91.5 | 5.5 | 471 | 1 | 5H2A_HUMAN |
| 19 | 90.5 | 5.5 | 456 | 1 | ENT2_HUMAN |
| 20 | 90.5 | 5.5 | 471 | 1 | 5H2A_MACMU |
| 21 | 89.5 | 5.4 | 471 | 1 | 5H2A_CRIGR |
| 22 | 89.5 | 5.4 | 471 | 1 | 5H2A_MOUSE |
| 23 | 89.5 | 5.4 | 471 | 1 | 5H2A_RAT |
| 24 | 89 | 5.4 | 347 | 1 | NU2M_DIDMA |
| 25 | 89 | 5.4 | 385 | 1 | YH7_SACDO |
| 26 | 89 | 5.4 | 429 | 1 | RNE_GUITH |
| 27 | 89 | 5.4 | 2232 | 1 | CCAE_RAT |
| 28 | 88.5 | 5.4 | 312 | 1 | OLP6_RAT |
| 29 | 88.5 | 5.4 | 318 | 1 | OLP6_CHICK |
| 30 | 88.5 | 5.4 | 433 | 1 | Y681_PASMU |
| 31 | 88.5 | 5.4 | 1882 | 1 | Y681_MYCPN |
| 32 | 87.5 | 5.3 | 821 | 1 | CAN3_RAT |
| 33 | 87 | 5.3 | 364 | 1 | ERG3_CANCA |

| | | | | | |
|----|------|-----|------|---|------------|
| 34 | 87 | 5.3 | 470 | 1 | 5H2A_PIG |
| 35 | 86.5 | 5.2 | 334 | 1 | YB77_METUA |
| 36 | 86.5 | 5.2 | 599 | 1 | DNAB_GUITH |
| 37 | 86.5 | 5.2 | 637 | 1 | NU5W_STRPU |
| 38 | 86.5 | 5.2 | 3655 | 1 | YAMB_SCHPO |
| 39 | 86 | 5.2 | 342 | 1 | PAFR_HUMAN |
| 40 | 85.5 | 5.2 | 312 | 1 | OLP6_CHICK |
| 41 | 85.5 | 5.2 | 583 | 1 | HASS_XENLA |
| 42 | 85.5 | 5.2 | 712 | 1 | FREE_YEAST |
| 43 | 85.5 | 5.2 | 821 | 1 | CAN3_MOUSE |
| 44 | 85.5 | 5.2 | 974 | 1 | ATXA_LEIDO |
| 45 | 85.5 | 5.2 | 974 | 1 | ATXB_LEIDO |

ALIGNMENTS

| RESULT | ID | Sequence | Standard | PRT | 315 AA |
|--------|--|----------|----------|-----|--------|
| 1 | RCE1_YEAST | 003530 | | | |
| AC | 003530 | | | | |
| DT | 01-NOV-1997 (Rel. 35, Created) | | | | |
| DT | 01-NOV-1997 (Rel. 35, Last sequence update) | | | | |
| DT | 20-AUG-2001 (Rel. 40, Last annotation update) | | | | |
| DE | CAAX PRENYL PROTEIN 2 (EC 3.4.22.-) (PRENYL PROTEIN-SPECIFIC | | | | |
| DE | ENDOPROTEASE 2) (PPSEP 2) (RMS AND A-FACTOR CONVERTING ENZYME) (RACE). | | | | |
| OS | RCE1 OR YMR274C OR YMR156.16C. | | | | |
| OS | Saccharomyces cerevisiae (Baker's yeast). | | | | |
| OC | Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; | | | | |
| OC | Saccharomycetales; Saccharomycetaceae; Saccharomyces. | | | | |
| OX | NCBI_TaxID:4932; | | | | |
| RN | [1] | | | | |
| RP | SEQUENCE FROM N.A. | | | | |
| RC | STRAIN-S288C / AB972; | | | | |
| RA | Lye G., Churcher C.M., Barrell B.G., Rajandream M.A.; | | | | |
| RL | Submitted (MAY-1995) to the EMBL/GenBank/DBJ databases. | | | | |
| RN | [2] | | | | |
| RP | FUNCTION. | | | | |
| RX | MEDLINE-97218305; PubMed-9065405; | | | | |
| RA | Boyartchuk V.L., Ashby M.N., Rine J.; | | | | |
| RT | "Modulation of Ras and a-factor function by carboxyl-terminal | | | | |
| RL | proteolysis."; | | | | |
| RL | Science 275:1796-1800(1997). | | | | |
| RN | [3] | | | | |
| RP | SUBCELLULAR LOCATION. | | | | |
| RX | MEDLINE-98409630; PubMed-9736709; | | | | |
| RA | Schmidt W.K., Tam A., Fujimura-Kamada K., Michaelis S.; | | | | |
| RT | "Endoplasmic reticulum membrane localization of Rce1p and Ste24p, | | | | |
| RL | yeast proteases involved in carboxyl-terminal CAAX protein processing | | | | |
| RL | and amino-terminal a-factor cleavage."; | | | | |
| RL | Proc. Natl. Acad. Sci. U.S.A. 95:11175-11180(1998). | | | | |
| RN | [4] | | | | |
| RP | CHARACTERIZATION. | | | | |
| RX | MEDLINE-20285442; PubMed-10825201; | | | | |
| RA | Trueblood C.E., Boyartchuk V.L., Picooglou E.A., Rozema D.; | | | | |
| RT | "The Caax proteases, Afc1p and Rce1p, have overlapping but distinct | | | | |
| RL | substrate specificities."; | | | | |
| RL | Mol. Cell. Biol. 20:4381-4392(2000). | | | | |
| CC | -1- FUNCTION: PROTEOLYTICALLY REMOVES THE C-TERMINAL THREE RESIDUES OF | | | | |
| CC | FARNESYLATED A-FACTOR MATING PHEROMONE AND OF RAS. | | | | |
| CC | -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. ENDOPLASMIC | | | | |
| CC | RETICULUM. | | | | |
| CC | -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C52. | | | | |
| CC | This SWISS-PROT entry is copyright. It is produced through a collaboration | | | | |
| CC | between the Swiss Institute of Bioinformatics and the EMBL outstation - | | | | |
| CC | the European Bioinformatics Institute. There are no restrictions on its | | | | |
| CC | use by non-profit institutions as long as its content is in no way | | | | |
| CC | modified and this statement is not removed. Usage by and for commercial | | | | |
| CC | entities requires a license agreement (See http://www.isb-sib.ch/announce/ | | | | |
| CC | or send an email to license@isb-sib.ch). | | | | |

| | | | |
|----|-----|---|-----|
| OY | 8 | LVLVLTSLSYPLPYATNSOPGSRKNDPRITKSMOKLTLMLSLNLFVPELOSQSSST | 67 |
| Dd | 25 | LAVLYGSSLYI---WSTRKH-----NRDHPITYVKRFASVSWMMLAAPFFVFESS--PELL | 75 |
| OY | 68 | SHISFKDAFTGLGITPGVYAALRPWFOFSQFVDLTACVAMLLTLCXGPVLDVFLYHLNLN | 127 |
| Dd | 76 | SRAVP-FKLGLLR-LEBLMOAWPIRY-----SLTVLLEFGPI-----FVN | 113 |
| OY | 128 | PKSIIIDEPHEPENTIW--SF-----RNFAPITEEIEITYSMLLTLYINLPHSOLSY | 179 |
| Dd | 114 | MQMSNVSTED--LDYRGSGSLTIWRNHVIAPLSEEVFRACAMP-----LIQSFSPL | 167 |
| OY | 180 | QQLFWPSLEFFGLAHANHAAYEOLOEGSMITVSIILTFCOYLITLTTEGGTLKEFYATGG | 239 |
| Dd | 168 | VAVEITP-LFEGVANHNIHIERLSIGVELSPALLI-GLQFIYTTLTGFEFSAFELPART-G | 224 |
| OY | 240 | NLMCCIIILHALCINTMGFP | 257 |
| Dd | 225 | HVMAPILVHAFCNNHGRL | 242 |

| RESULT | 3 | |
|--------|--|------------------------|
| ID | FAC2_MOUSE | |
| AC | P57791; 09PE68; | STANDARD; PRT; 329 AA. |
| DT | 20-AUG-2001 (Rel. 40, Created) | |
| DT | 20-AUG-2001 (Rel. 40, Last sequence update) | |
| DT | 20-AUG-2001 (Rel. 40, Last annotation update) | |
| DE | CAAX PREN1L PROTEASE 2 (EC 3.4.22.-) (PREN1L PROTEIN-SPECIFIC ENDOPEPTIDASE 2) (FARNESYLATED-PROTEINS CONVERTING ENZYME 2) (FACE-2). | |
| OS | MUS MUSCULUS (Mouse). | |
| OC | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; | |
| OC | Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. | |
| OX | NCBI_Taxid:10090; | |
| RN | [1] | |
| RP | SEQUENCE FROM N.A. | |
| RA | Cadinance J., Freije J.M.P.; | |
| RT | "Characterization and expression analysis of the gene encoding the murine Caax protease face-2."; | |
| RL | Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases. | |
| RN | [2] | |
| RP | KNOCK-OUT. | |
| RX | MEDLINE-99185053; PubMed-10085069; | |
| RA | Kim E., Ambrozic P., Otto J.C., Taylor B., Ashby M., Shannon K., Casey P.J., Young S.G.; | |
| RT | "Disruption of the mouse Rce1 gene results in defective Ras processing and mislocalization of Ras within cells."; | |
| RL | J. Biol. Chem. 274:8383-8390(1999). | |
| CC | -1- FUNCTION: PROTOPLASMICALLY REMOVES THE C-TERMINAL THREE RESIDUES OF FARNESYLATED AND GERANYLATED PROTEINS. SEEMS TO BE ABLE TO PROCESS K-RAS, N-RAS, H-RAS, RAP1B AND G-GAMMA-1 (BY SIMILARITY). | |
| CC | -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. ENDOPLASMIC RETICULUM (BY SIMILARITY). | |
| CC | -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C52. | |
| CC | ----- | |
| CC | This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch). | |
| CC | ----- | |
| DR | EMBL; AJ251644; CAC17013.1; - | |
| DR | EMBL; AJ251645; CAC17014.1; - | |
| DR | MEROPS; C52.002; - | |
| DR | MCD; MG1:1336895; Rce1. | |
| KM | Hydrolase; Thiol protease; Transmembrane; Endoplasmic reticulum. | |
| FT | TRANSMEM 36 56 | POTENTIAL. |
| FT | TRANSMEM 75 95 | POTENTIAL. |
| FT | TRANSMEM 112 132 | POTENTIAL. |
| FT | TRANSMEM 193 213 | POTENTIAL. |

| FT | TRANSMEM | 226 | 246 | POTENTIAL. |
|--|----------|---------------------------------------|-------------------------|------------------------|
| FT <th>TRANSMEM</th> <td>254 <td>274 <th>POTENTIAL.</th> </td></td> | TRANSMEM | 254 <td>274 <th>POTENTIAL.</th> </td> | 274 <th>POTENTIAL.</th> | POTENTIAL. |
| FT <th>TRANSMEM</th> <td>282 <td>302 <th>POTENTIAL.</th> </td></td> | TRANSMEM | 282 <td>302 <th>POTENTIAL.</th> </td> | 302 <th>POTENTIAL.</th> | POTENTIAL. |
| SO <td>SEQUENCE</td> <td>329 AA;</td> <td>35866 MM;</td> <td>C2A49617EBC77EC8 CRC64</td> | SEQUENCE | 329 AA; | 35866 MM; | C2A49617EBC77EC8 CRC64 |

| | | | | |
|--------------------------|--------|--------------------|------------|-------------|
| Query Match | 14.08; | Score 230.5; | DB 1; | Length 329; |
| Best Local Similarity | 30.68; | Pred. No. 2.6e-11; | | |
| Matches 85; Conservative | 40; | Mismatches 82; | Indels 71; | Gaps 16 |

| | | | |
|----|-----|---|-----|
| QY | 4 | ESTPLVLLTYSIVELPLVLTATSOEGRSKRDPRIKRMOKLTMTLSINF-LVPLFQSQ | 62 |
| Db | 40 | FSCF-----SLACSVSGSLYWKSE--LPRDHPAIKRR--STSLVYVSSISPLCYLLMRE | 91 |
| QY | 63 | LSST---TSHISEKDAFLGLGIIPGYAALRNPMQSFQWVKDLTKCYAMLLT--LYCGPV | 117 |
| Db | 92 | LTGIQPTSLTLLM-GRLEGIFPA--ALLP-----LLTMTLLEFGPL | 131 |
| QY | 118 | LDPV-----LYNLLMPKKSILIEDFNHEFLNMS-----FRNEFPAPTEEIF | 139 |
| Db | 132 | MQLSMDCRDLTDGLKAVLLAPRS-----MARCLSTMRYLRNOVIAPLDELV | 178 |
| QY | 160 | YTSMLTTYLNLRPHSOLSYQOLEMORSLFFGLANAHAYEOLQEGSMTYSILLTTCFQ | 219 |
| Db | 179 | FRACMLRM--LACSTGLG-PAVETCP-LREGVANHNIIEQLRFQSSGSIIVSAFQ | 233 |
| QY | 220 | ILYTLTEGGLTKRVFVNTGGMSCILLALNCINMGFR | 257 |
| Db | 234 | ESYAVFGAATALELIFT-GHLIGRPVLCSSCNMGFR | 270 |

| | | | | |
|----|------------|---|-----------------------------------|--------------|
| | RESULT | 4 | | |
| CC | FAC2_HUMAN | | | |
| CC | ID | PAC2_HUMAN | STANDARD; | PRT; 329 AA. |
| CC | AC | Q9Y2S6; | | |
| CC | DT | 20-AUG-2001 | (Rel. 40, Created) | |
| CC | DT | 20-AUG-2001 | (Rel. 40, Last sequence update) | |
| CC | DT | 20-AUG-2001 | (Rel. 40, Last annotation update) | |
| CC | DE | CXAX PRENYL PROTEASE 2 (EC 3.4.22.-) (PRENYL PROTEIN-SPECIFIC ENDOPROTEASE 2) (FARNESYLATED-PROTEINS CONVERTING ENZYME 2) (FACE-2). | | |
| CC | GN | RCELA OR RCEL OR FACE2. | | |
| CC | OS | Homo sapiens (Human). | | |
| CC | OC | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; | | |
| CC | OC | Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. | | |
| CC | OX | NCBI_Taxid=9606; | | |
| CC | RN | [1] | | |
| CC | RP | SEQUENCE FROM N.A. | | |
| CC | RC | TISSUE=Ovary; | | |
| CC | RX | MEDLINE=99303558; PubMed=10373325; | | |
| CC | RA | Frelle J.M.P., Blay P., Pendas A.M., Caninaos J., Crespo P., | | |
| CC | RA | Lopez-Otin C.; | | |
| CC | RT | "Identification and chromosomal location of two human genes encoding enzymes potentially involved in proteolytic maturation of farnesylated proteins."; | | |
| CC | RL | Genomics 58:270-280(1999). | | |
| CC | RN | [2] | | |
| CC | RP | SEQUENCE FROM N.A. | | |
| CC | RX | MEDLINE=99185052; PubMed=10085066; | | |
| CC | RA | Otto J.C., Kim E., Young S.G., Casey P.J.; | | |
| CC | RT | "Cloning and characterization of a mammalian ptenyl protein-specific protease."; | | |
| CC | RL | J. Biol. Chem. 274:8379-8382(1999). | | |
| CC | -1- | FUNCTION: PROTEOLYTICALLY REMOVES THE C-TERMINAL THREE RESIDUES OF FARNESYLATED AND GERANYLATED PROTEINS. SEEMS TO BE ABLE TO PROCESS K-RAS, N-RAS, H-RAS, RAP1B AND G-GAMMA-1. | | |
| CC | -1- | SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. ENDOPLASMIC RETICULUM (BY SIMILARITY). | | |
| CC | -1- | TISSUE SPECIFICITY: UBIQUITOUS. | | |
| CC | -1- | SIMILARITY: BELONGS TO PEPTIDASE FAMILY C52. | | |
| CC | -1- | THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outpost - its European Bioinformatics Institute. There are no restrictions on its | | |

CC entities requires a license agreement (See [http://www.isb-sib.ch/announcement/](http://www.isb-sib.ch/announcement)
CC or send an email to license@isb-sib.ch).

DR EMBL: AF081829; AAC05526.1; -.
DR InterPro: IPR003918; NMDHnb_oxidrcdsc4.
DR InterPro: IPR001750; Oxidored_q1.
DR Pfam: PF00361; Oxidored_q1.1.
KW Oxidoreductase; NAD; Ubiquinone; Mitochondrion.
SQ SEQUENCE 433 AA; 50758 MW; 936638781633CD04 CRC64

| | | | | |
|-----------------------|------------------|-----------------|------------|-------------|
| Query Match | 6.1%; | Score 100.5; | DB 1; | Length 433; |
| Best Local Similarity | 22.7%; | Pred. No. 0.61; | | |
| Matches 61; | Conservative 37; | Mismatches 88; | Indels 83; | Gaps 14; |

```

QY 1 MGOSTFVLLVYIISVLYPLVANSOEGSKRNDPRITKSMOKLTIMLSLFLVPLQ 60
   :: : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 50 LMSLSMVLIVYISILMIM-----AKSNM--NYKNKINFFYLLIMNINLFICFM- 97
   :: : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 61 SOLSSTTSHISFKDAFL-----GLGIIP----GYVALPMPWQSFVKDLTCVA 107
   :: : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 98 -LENLIMFYLFPEAVLFPIILMISGWSOPERIQAGFY----- 134
   :: : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 108 MLTLVLCGPVLDFVLXHLNKRSSILDFYHFNFINISSFRNFIAPITBELFYTSMLT 164
   :: : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 135 MLMTTVGSLPLILMLIKNOSLIT-----FNEWLFENMGF-----IFF--LMLIG 179
   :: : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 168 YLNIIPHSLSYQOLF--WPSLFEGLAHAAHAYEOLOEGSMTVSILTLTTCFOIL- 222
   :: : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 180 FLVRIKIP-----MFLHLMIPK-----AHV-----EAPRAGSMILAGVLKIGFYGLRFK 224
   :: : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 223 TTLFGGLTKFYFVPTGMLMCCIIILHNLG 251
   :: : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 225 SFPLDLTKESFVLIVISMGGVLAITISFC 253
   :: : : : : : : : : : : : : : : : : : : : : : : : : : :

```

| RESULT | 7 | | | |
|--------|--|-----------------------------------|---------------|--|
| ID | MRAY_STRAU | STANDARD | PRT | 321 AA. |
| AC | 007322 | 024815 | | |
| DT | 30-MAY-2000 | (Rel. 39, Created) | | |
| DT | 30-MAY-2000 | (Rel. 39, Last sequence update) | | |
| DT | 20-AUG-2001 | (Rel. 40, Last annotation update) | | |
| DE | PHOSPHO-N-ACETYLTRANSFERASE | PHOSPHOTRANSFERASE | (EC 2.7.8.13) | (UDP-MUNAC-PENTAPEPTIDE PHOSPHOTRANSFERASE). |
| GN | MRAY. | | | |
| OS | Staphylococcus aureus. | | | |
| OC | Bacteria; Firmicutes; Bacillus/Clostridium group; | | | |
| OC | Bacillus/Staphylococcus group; Staphylococcus. | | | |
| OX | NCBI_TaxId=1280; | | | |
| RN | [1] | | | |
| RN | SEQUENCE FROM N.A. | | | |
| RP | STRAIN=ATCC 8325-4; | | | |
| RX | MEDLINE=97431524; PubMed=9287029; | | | |
| RA | Pucci M.J., Thanassi J.A., Discotto L.F., Kessler R.E., | | | |
| RA | Dougherty T.J.; | | | |
| RT | "Identification and characterization of cell wall-cell division gene | | | |
| RT | clusters in pathogenic gram-positive cocci."; | | | |
| J. | Bacteriol. 179:5632-5635(1997). | | | |
| RN | [2] | | | |
| RP | SEQUENCE FROM N.A. | | | |
| RC | STRAIN=NCTC 8325; | | | |
| RX | MEDLINE=98241542; PubMed=9573165; | | | |
| RA | Wada A., Watanabe H.; | | | |
| RT | "Penicillin-binding protein 1 of Staphylococcus aureus is essential | | | |
| RT | for growth."; | | | |
| J. | Bacteriol. 180:2759-2765(1998). | | | |
| RN | [3] | | | |
| RP | SEQUENCE FROM N.A. | | | |
| RC | STRAIN=R27; | | | |
| RA | El-Sherbini M., Geissler W.M., Azzolina B.A., Yuan X., Hyland S.A., | | | |
| RA | Anderson M.S., Pompiliano D.L.; | | | |
| RT | "Cloning and characterization of the Staphylococcus aureus mray gene | | | |

RT encoding phospho-N-acetylmutamoyl-pentapeptide translocase." ,
RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.

RX MEDLINE-20032370; PubMed-10564498.
 RA Bouches A., Mengin-Lecreux D., Le Beller D., Van Heijenoort J.;
 RT "Topological analysis of the may protein catalyzing the first
 membrane step of peptidoglycan synthesis.";
 RL Mol. Microbiol. 34:576-585(1999).

CC -I- FUNCTION: FIRST STEP OF THE LIPID CYCLE REACTIONS IN THE BIOSYNTHESIS OF THE CELL WALL PEPTIDOGLYCAN (BY SIMILARITY). CC -I- CATALYTIC ACTIVITY: UDP-N-ACETYLGLUCOSAMINE-1-ALANINE-D-GLUTAMYL-L-LYSYL-D-ALANINE + UNDECAPRENYL PHOSPHATE = UMP + N-ACETYLGLUCOSAMINE-1-ALANINE-D-GLUTAMYL-L-LYSYL-D-ALANINE-D-ALANINE.

CC -1- PATHWAY: PEPTIDOLYCAN BIOSYNTHESIS.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -1- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 4. MRAV
CC SUBFAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed, usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement>
CC or send an email to license@isb-sib.ch).

CC -----
 DR EMBL; U94706; AAC45625.1; -
 DR EMBL; AB007500; BAA2555.1; -
 DR EMBL; AF034153; AAD01977.1; -
 DR InterPro: IPR000715; Glycos.transf.4.
 DR Pfam: PF00953; Glycos.transf.4; 1.
 DR PROSITE; PS01347; MRA1_1; 1.
 DR PROSITE; PS01348; MRA1_2; 1.
 DR Peptidoglycan synthetase; Cell division; Transferase; Transmembrane

| | | | | | |
|----|----------|-----|-----|--|--|
| FT | TRANSMEM | 1 | 21 | | |
| FT | DOMAIN | 22 | 52 | CYTOPLASMIC (POTENTIAL) . | |
| FT | TRANSMEM | 53 | 69 | | |
| FT | DOMAIN | 70 | 75 | EXTRACELLULAR (POTENTIAL) . | |
| FT | TRANSMEM | 76 | 91 | | |
| FT | DOMAIN | 92 | 111 | CYTOPLASMIC (POTENTIAL) . | |
| FT | TRANSMEM | 112 | 133 | | |
| FT | DOMAIN | 134 | 150 | EXTRACELLULAR (POTENTIAL) . | |
| FT | TRANSMEM | 151 | 165 | | |
| FT | DOMAIN | 166 | 176 | CYTOPLASMIC (POTENTIAL) . | |
| FT | TRANSMEM | 177 | 197 | | |
| FT | DOMAIN | 198 | 202 | EXTRACELLULAR (POTENTIAL) . | |
| FT | TRANSMEM | 203 | 216 | | |
| FT | DOMAIN | 217 | 232 | CYTOPLASMIC (POTENTIAL) . | |
| FT | TRANSMEM | 233 | 246 | | |
| FT | DOMAIN | 247 | 249 | EXTRACELLULAR (POTENTIAL) . | |
| FT | TRANSMEM | 250 | 268 | | |
| FT | DOMAIN | 269 | 300 | CYTOPLASMIC (POTENTIAL) . | |
| FT | TRANSMEM | 301 | 320 | | |
| FT | DOMAIN | 321 | 321 | EXTRACELLULAR (POTENTIAL) . | |
| FT | CONFLICT | 104 | 104 | Q -> P (1N REF. 1) . | |
| FT | CONFLICT | 254 | 254 | Q -> I (1N REF. 1) . | |
| FT | CONFLICT | 299 | 321 | KVYTFWAGVGLGIGTWIGVH -> ESTYSILGCSDFRE | |

| | | | | |
|-----------------------|------------------|------------------|------------|-------------|
| Query Match | 6.08; | Score 99; | DB 1; | Length 321; |
| Best Local Similarity | 21.28; | Pred. No. 0.586; | | |
| Matches 69; | Conservative 62; | Mismatches 118; | Indels 76; | Gaps 16; |

QY 8 LVLLXLSLSTYPLPLX-----ATSQPESGRKNDNRRTIKSRKQKTLINL-----ISLFLVLPF 58
 10 LVTFVFLVPLVPLIPLRKRMKFSQSIRESRQPSHMKTSTPTMGGLFTLLSLIVTSIAVIF 69
 QY 59 LOSQLSTSTHSISKDAFLGL-GIIPGYAALPNPWFQSFVKDLTCVAMLLTYGC-P 116

Db 70 VDO--ANDIILLFVITGIGFIDDIYK--KNOGLTSKOKELAQI-----GIA 119
 QY 117 VLDPVL---YHLNPKSSILEDPEHEPLNI---MSRNFIPAPITEEFYTSMLTTYLN 170
 Db 120 IIFVLSNVHLVNFSTSI-----HIPFTNVAIPISFAVIT-----157
 QY 171 LIPHSLSYQOLFQPSLFFGLAHNAHAYEOLQSGSMVTYSILTTCPQILYTLFGGLT 230
 Db 158 -----VFMO-----VGFSSNAVNLTDGL-DGLATGLSIIIGFTWYAIM-----192
 QY 231 KFEVVRGNGNLMCCIIILHALCNMGFP-GPSRLNHLFVYVKKAGRISKLVSTNNKCYFA 289
 Db 193 SFVAGETAIGIFCIIIMFALLGFLPYNNIPAKVEMGDT-GSLALGIFATISIMLNDELS 251
 QY 290 LVLGLISLKDITLQTLVCTPGYRIT 314
 Db 252 LIFGLVFEVITLSVLMQVASFRLT 276

RESULT 8
 ID ATP6_ANOGA STANDARD: PRT: 226 AA.
 AC P34834;
 DT 01-FEB-1994 (Rel. 28, Created)
 DT 01-FEB-1994 (Rel. 28, Last sequence update)
 DT 15-JUL-1998 (Rel. 36, Last annotation update)
 DE ATP SYNTHASE A CHAIN (EC 3.6.1.34) (PROTEIN 6).
 OS Anopheles gambiae (African malaria mosquito).
 OC Mitochondrion.
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae;
 OC Anopheles.
 OX NCBI_TaxID=7165;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=G3;
 RX MEDLINE=97242550; PubMed=9087549;
 RA Beard C.B., Hamm D.M., Collins F.H.;
 RT "The mitochondrial genome of the mosquito Anopheles gambiae: DNA
 sequence, genome organization, and comparisons with mitochondrial
 sequences of other insects.";
 RL Insect Mol. Biol. 2:103-104(1993).
 CC -1- FUNCTION: KEY COMPONENT OF THE PROTON CHANNEL. IT MAY PLAY A
 DIRECT ROLE IN THE TRANSLLOCATION OF PROTONS ACROSS THE MEMBRANE.
 CC -1- SUBUNIT: F-TYPE ATPASES HAVE 2 COMPONENTS, CF(1) - THE CATALYTIC
 CORE - AND CF(0) - THE MEMBRANE PROTON CHANNEL. CF(1) HAS FIVE
 SUBUNITS: ALPHA(3), BETA(3), GAMMA(1), DELTA(1), EPSILON(1). CF(0)
 HAS THREE MAIN SUBUNITS: A, B AND C.
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
 CC -1- SIMILARITY: BELONGS TO THE ATPASE A CHAIN FAMILY.
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: L20934; AAD12194.1;
 DR InterPro: IPR000568; ATP_synA.
 DR Pfam: PF00119; ATP-syntA.1.
 DR PRINTS: PR00123; ATPASEA.
 DR PROSITE: PS00449; ATPASE_A.1.
 KW Hydrogen ion transport; CF(0); Mitochondrion; Transmembrane.
 SQ SEQUENCE 226 AA; 25276 MW; 0AE0AB8BA78AABAD CRC64;

Query Match 5.8%; Score 96.5; DB 1; Length 226;
 Best Local Similarity 23.4%; Pred. No. 0.63;
 Matches 59; Conservative 38; Mismatches 98; Indels 57; Gaps 11;

QY 49 LISNLFVPELOSLSSTSHSEK--DAFLGIIIPGYAALPNPQ--FSQVFKDLTK 104
 Db 1 MMTNLFV-----PDPSTIILNLSNMLSTFGLFLIVSYVLMNRRQVIMNNLLTLHK 56
 QY 105 CVAMNL--TLXCGVLDV-LYHLLNPKSSILEDPEHEPLNI---MSRNFIPAPITEEFY 161
 Db 57 EFKTLGSPGSHNGSTLMEISLFSLI-----MENNFLGLFPY-----1FTS 96
 QY 162 SMLITTYNLNLPHSLSYQOLFQPSLFFGLAHNAHAYEOLQSGSMVTYSILTTCPQIL 221
 Db 97 TSHLTLLALFPLMLSLMXYG-----INRQHEFAHLVPGCTPPVLMFPVVCJETI 149
 QY 222 YTTLFGGLTKEVFRGTGNLMCCIIILHALCNMGFPGPSRLNHLFVYVKKAGRISKLV 281
 Db 150 SNVIRPGF---LAVRLTANM---TAGHLLTLTLGNTGPMASNY-----LILS 190
 QY 282 IWNKCYFALLV 293
 Db 191 LILTTQIALLV 202

RESULT 9
 ID Y255_MYCPN STANDARD: PRT: 534 AA.
 AC P75422;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE HYPOTHETICAL PROTEIN MG255 HOMOLOG (H91_08F534).
 GN MP358 OR MP478.
 OS Mycoplasma pneumoniae.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
 OC Mycoplasmataceae; Mycoplasma.
 OX NCBI_TaxID=2104;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 29342 / M129;
 RX MEDLINE=97105885; PubMed=8948633;
 RA Himmelfreich R., Hilbert H., Plagens H., Pirkel E., Li B.-C.,
 RA Hermann R.;
 RT "Complete sequence analysis of the genome of the bacterium Mycoplasma
 pneumoniae.";
 RL Nucleic Acids Res. 24:4420-4449(1996).
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AE000047; AAB96126.1;
 KW Hypothetical protein; Transmembrane; Complete proteome.
 FT TRANSMEM 149 169 POTENTIAL.
 FT TRANSMEM 185 205 POTENTIAL.
 SQ SEQUENCE 534 AA; 62643 MW; 08B95Cd4515FA66C CRC64;

Query Match 5.8%; Score 96.5; DB 1; Length 534;
 Best Local Similarity 26.3%; Pred. No. 1.6;
 Matches 56; Conservative 35; Mismatches 85; Indels 37; Gaps 11;
 QY 3 QFSTFVLVLYISISYVLPVATSOPEGSKRDNPRTIKSRMOKTILISNLFVPLQSQ 62
 Db 270 QAMNLLLLKQTNCFVL-----VCCKESNMILLINKVKEPDINKOSSLYL--DKSQ 319
 QY 63 LSSITSHISFKDAFLGIGI-----IPGYAALPNPQFSQF---VVDLTRCVAMLLTLY 113
 Db 320 ISPLAQISKYNLLFEELALADMFYLEDFFALLKTPQIVNLFERRIKNLKEFHQPTQL 379
 QY 114 CGPVLDFV-LYHLLNPKSSILEDPEHEPLNI---MSRNFIPAPITEEFYTSMLTTYLN 170

[illegible]

```

Matches 37: Conservative 31: Mismatches 36: Indels 48: Gaps 9:
QY 7 FLVLLYIIS-VYLEPIVTSOPEGSKRDNPRTIKSRMOKLITMILSNFL-----VFPL 59
Db 196 FLIPLEFIMIGCYLVLIHMLHGRTRK-----LKEKVEKSIIRIITLTVGLVCFMCP- 248
QY 60 OSQSLSTSHISFKNAFLGLIGIPGYAALPNP--QFQPVKDLKCVAMLLTLCGPVL 118
Db 249 -----HICE-AFLMLGGENSY-----NPGCAFTFLMLNSTC-----L 281
Db 119 DFVLYHLNPKRSILEDIFYHEFLNIWSPRNF 150
Db 282 DVLILYIYS-----KQFQARIYSMLLRNVL 307

RESULT 11
YA35--SCHPO STANDARD: PRT: 426 AA.
ID YA35_SCHPO 009712:
AC 009712:
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE HYPOTHETICAL 48.8 KDA PROTEIN C18B11.05 IN CHROMOSOME 1.
GN SPAC18B11.05.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomyces.
OX NCBI_TaxID=4896;
RN 11
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RA Devlin K., Churcher C.M., Barrell B.G., Rajandream M.A., Walsb S.V.:
RA Submitted (Aug-1995) to the EMBL/Genbank/DBJ databases.
CC 1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC 1- SIMILARITY: TO YEAST YBR004C AND SOME, TO C.ELEGANS ZK1321.3.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sdb.ch/announce/
CC or send an email to license@isb-sdb.ch).
CC -----
CC EMBL: Z50728; CAA90590.1; -.
DR Hypothetical protein: Transmembrane.
KW Hypothetical protein: Transmembrane.
FT TRANSMEM 33 53
FT TRANSMEM 104 124 POTENTIAL.
FT TRANSMEM 129 149 POTENTIAL.
FT TRANSMEM 182 202 POTENTIAL.
FT TRANSMEM 223 243 POTENTIAL.
FT TRANSMEM 264 284 POTENTIAL.
FT TRANSMEM 319 339 POTENTIAL.
FT TRANSMEM 351 371 POTENTIAL.
FT TRANSMEM 403 423 POTENTIAL.
FT SEQUENCE 426 AA; 48824 MW; CAACF57FAD4E0FC CRC64;
SQ
Query Match 5.8%; Score 96; DB 1; Length 426;
Best Local Similarity 21.9%; Pred. No. 1.4;
Matches 58: Conservative 40: Mismatches 93: Indels 74: Gaps 12:
QY 11 LYSISVLPIATATSQPEGSKRDNPRTIKSRMOKLITMILSNFLVLPFLOSLSSTSHI 70
Db 179 IYMSGYEESLFLAASFLG-----LLFLPKQOYPAALFMSLATIRNSG 223
QY 71 SFKDAFLGLGIPGYAALPNP---WQSFQVNDLTKCVAMLLTVGCPVLDVFLY--- 123
Db 224 IFWCIFGMPALGTLKISIERQLQTFPMQVSGVGTGKCLITLVDFPNQYVGFLFCPG 283
QY 124 -----HLNPKSKSILEDPIH-EFLNIWSPR---NPIFA-----PITTEIFY--TSM 164

```

```

Db 284 VAMCKSLPIYP--AVOEKYMVGFRLRYMTLNIPNLFALLSIIPILFALFYSISGST 341
OY 165 LTTYNLPHSQSLYSQOLFWQPSLEFGLAHAAHEQLOEGSMPTVSLITTCPIILTT 224
Db 342 LHSFSTKSHLYLS--ALY---LYIGCFHHM-----TOVLNRMSSALPLLYWS 384
OY 225 L-----FGLTKFEV 235
Db 385 MAHATLVAKSRNLKAFGHCILFWMI 409

RESULT 12
ATP6_DROSI STANDARD: PRT: 224 AA.
ID ATP6_DROSI
AC P50269:
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE ATP SYNTHASE A CHAIN (EC 3.6.1.34) (PROTEIN 6).
GN ATP6 OR ATPASE6.
OS Drosophila simulans (Fruit fly).
OC Mitochondrion.
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7240;
RN [1]
RP MEDLINE=93374296; PubMed=8365657;
RA Kaneko M., Satta Y., Matsuura E.T., Chigusa S.I.;
RT "Evolution of the mitochondrial ATPase 6 gene in Drosophila:
RT unusually high level of polymorphism in D. melanogaster.";
RL Genet. Res. 61:195-204(1993).
CC -1- FUNCTION: KEY COMPONENT OF THE PROTON CHANNEL. IT MAY PLAY A
CC DIRECT ROLE IN THE TRANSLOCATION OF PROTONS ACROSS THE MEMBRANE.
CC -1- SUBUNIT: F-TYPE ATPASES HAVE 2 COMPONENTS, CF(1) - THE CATALYTIC
CC CORE - AND CF(0) - THE MEMBRANE PROTON CHANNEL. CF(1) HAS FIVE
CC SUBUNITS: ALPHA(3), BETA(3), GAMMA(1), DELTA(1), EPSILON(1). CF(0)
CC HAS THREE MAIN SUBUNITS: A, B AND C.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -1- SIMILARITY: BELONGS TO THE ATPASE A CHAIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: S649377; AAD13958.1;
DR FlyBase: FBgn0012875; Dsln\mt:ATPase6.
DR InterPro: IPR000568; ATP_synth_A.
DR Pfam: PF00119; ATP-synt_A.1.
DR PRINTS: PR00123; ATPASEA.
DR PROSITE: PS00449; ATPASEA.1.
KW Hydrogen ion transport; CF(0); Mitochondrion; Transmembrane.
SQ SEQUENCE 224 AA; 25065 MW; 9480FE8215156782 CRC64;

Query Match 5.88; Score 95.5; DB 1; Length 224;
Best Local Similarly 24.08; Pred. No. 0.75;
Matches 61; Conservative 40; Mismatches 90; Indels 63; Gaps 15;

OY 49 LISNLFVLPFLQSLSTSHSKDAFLGIGITPGYALPN---PWOSQVFKDLTK 104
Db 1 MMTNLFVS-FDPSAIFENS--LNMWSTFLGLMIPSIYMWPSRYNIWN----- 47
OY 105 CVAMLLTYCGPVLDFVHLLNPKSSILEDVFEHLNIMS---FRNFIPATEEIPYT 161
Db 48 --SILLTIH---KEFKT--LLGPSGHNGSTFT--FISLEFLILFNMM--GLPPIYITS 95
OY 162 SMLTTYLNL-IPHSQSLYSQOLFWQPSLEFG-LAAHHAHYEOLQEGSMPTVSLITTCFQ 219

```

```

Db 96 TSHLTLTSLALP-----LMICFMYLXGWINNTQGMFHVYQGNPALIMPRWCIE 146
OY 220 ILTYTLFEGGLTKFEVFTGNCIMCCIIILHALCNIMGPGPSRLNHTFVVDKAGRISK 279
Db 147 TISNIIRPGT---LAVRLTANM---IAGHLTLTLTGNGPSMSYL-----L 186
OY 280 VSIMNKCYFALLVL 293
Db 187 VFLTLTAQIALLVL 200

```

```

RESULT 13
ATP6_DROYA STANDARD: PRT: 224 AA.
ID ATP6_DROYA
AC P00851:
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE ATP SYNTHASE A CHAIN (EC 3.6.1.34) (PROTEIN 6).
GN ATP6 OR ATPASE6.
OS Drosophila yakuba (Fruit fly).
OC Mitochondrion.
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7245;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=2317.6 IVORY COAST;
RA MEDLINE=86089137; PubMed=3001325;
RX Clary D.O., Wolstenholme D.R.;
RT "The mitochondrial DNA molecular of Drosophila yakuba: nucleotide
RT sequence, gene organization, and genetic code.";
RL J. Mol. Evol. 22:252-271(1985).
RN [2]
RP SEQUENCE FROM N.A.
RC MEDLINE=83246544; PubMed=6306579;
RX Clary D.O., Wolstenholme D.R.;
RT "Nucleotide sequence of a segment of Drosophila mitochondrial DNA
RT that contains the genes for cytochrome c oxidase subunits II and III
RT and ATPase subunit 6.";
RL Nucleic Acids Res. 11:4211-4227(1983).
CC -1- FUNCTION: KEY COMPONENT OF THE PROTON CHANNEL. IT MAY PLAY A
CC DIRECT ROLE IN THE TRANSLOCATION OF PROTONS ACROSS THE MEMBRANE.
CC -1- SUBUNIT: F-TYPE ATPASES HAVE 2 COMPONENTS, CF(1) - THE CATALYTIC
CC CORE - AND CF(0) - THE MEMBRANE PROTON CHANNEL. CF(1) HAS FIVE
CC SUBUNITS: ALPHA(3), BETA(3), GAMMA(1), DELTA(1), EPSILON(1). CF(0)
CC HAS THREE MAIN SUBUNITS: A, B AND C.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -1- SIMILARITY: BELONGS TO THE ATPASE A CHAIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X03240; CAA26989.1;
DR EMBL: X00924; CAA25442.1;
DR PIR: G34477; PMFE6Y.
DR PIR: E25797; E25797.
DR FlyBase: FBgn0013177; Dyak\mt:ATPase6.
DR InterPro: IPR000568; ATP_synth_A.
DR Pfam: PF00119; ATP-synt_A.1.
DR PRINTS: PR00123; ATPASEA.
DR PROSITE: PS00449; ATPASEA.1.
KW Hydrogen ion transport; CF(0); Mitochondrion; Transmembrane.
SQ SEQUENCE 224 AA; 25141 MW; 27D6B8F5791FDD24 CRC64;

```

| FT | TRANSMEM | 101 | 120 | 3 | POTENTIAL |
|----|----------|---------|-----------|----------------------|---------------------------|
| FT | DOMAIN | 121 | 139 | 4 | CYTOPLASMIC (POTENTIAL) |
| FT | TRANSMEM | 140 | 164 | 4 | (POTENTIAL) |
| FT | TRANSMEM | 165 | 205 | 5 | EXTRACELLULAR (POTENTIAL) |
| FT | TRANSMEM | 206 | 226 | 5 | (POTENTIAL) |
| FT | DOMAIN | 227 | 239 | 6 | CYTOPLASMIC (POTENTIAL) |
| FT | TRANSMEM | 240 | 260 | 6 | (POTENTIAL) |
| FT | DOMAIN | 261 | 271 | 7 | EXTRACELLULAR (POTENTIAL) |
| FT | TRANSMEM | 272 | 292 | 7 | (POTENTIAL) |
| FT | DOMAIN | 293 | 312 | 7 | CYTOPLASMIC (POTENTIAL) |
| FT | CARBOHYD | 5 | 5 | N-LINKED (GLCNAC. .) | (POTENTIAL) |
| FT | DISULFID | 97 | 179 | BY SIMILARITY | |
| SO | SEQUENCE | 312 AA; | 35179 MW; | 56D164BFP9D59EA9 | CNC64; |

| Query Match | Best Local Similarity | 5.7%; | Score 94.5; | DB 1; | length 312; |
|-------------|-----------------------|--------------|-------------|------------|-------------|
| Matches | 60; | Conservative | 55; | Mismatches | 97; |
| | | | | Indels | 11; |
| | | | | Gaps | 16; |

| QY | 46 | TIMLISNLFVLPFLOSQSSTT-----SHISFKAFGLGILIPGYAALPMPQSPQFV | 99 |
|---|--|--|-----|
| DB <td>36<td>TITLTNTGLIALISIDLOQPMYIFLONSFTDA-----VSVIVIPKMLATFL<td>86</td></td></td> | 36 <td>TITLTNTGLIALISIDLOQPMYIFLONSFTDA-----VSVIVIPKMLATFL<td>86</td></td> | TITLTNTGLIALISIDLOQPMYIFLONSFTDA-----VSVIVIPKMLATFL <td>86</td> | 86 |
| QY <td>100<td>KDLTFCAVMLLTVCGPVLDPVLYHLLMPKSSILDEYHEFLINMSFRNFIAPTEEIF<td>159</td></td></td> | 100 <td>KDLTFCAVMLLTVCGPVLDPVLYHLLMPKSSILDEYHEFLINMSFRNFIAPTEEIF<td>159</td></td> | KDLTFCAVMLLTVCGPVLDPVLYHLLMPKSSILDEYHEFLINMSFRNFIAPTEEIF <td>159</td> | 159 |
| DB <td>87<td>EE-TKTTIS-----YVGCILQYFSEVLLVRECLL-----LAVAYDR--YAAICKPPL<td>131</td></td></td> | 87 <td>EE-TKTTIS-----YVGCILQYFSEVLLVRECLL-----LAVAYDR--YAAICKPPL<td>131</td></td> | EE-TKTTIS-----YVGCILQYFSEVLLVRECLL-----LAVAYDR--YAAICKPPL <td>131</td> | 131 |
| QY <td>160<td>YTSML-----TYTLNDLPHS-----QLSY-----QDLFWQPSLFGJLAH<td>196</td></td></td> | 160 <td>YTSML-----TYTLNDLPHS-----QLSY-----QDLFWQPSLFGJLAH<td>196</td></td> | YTSML-----TYTLNDLPHS-----QLSY-----QDLFWQPSLFGJLAH <td>196</td> | 196 |
| DB <td>132<td>YPAIMTKAVCMRLVKGXLSIAFLNPLVHTSGGLKLSFSCSNVNVNPFQCNPSPLQISSS<td>191</td></td></td> | 132 <td>YPAIMTKAVCMRLVKGXLSIAFLNPLVHTSGGLKLSFSCSNVNVNPFQCNPSPLQISSS<td>191</td></td> | YPAIMTKAVCMRLVKGXLSIAFLNPLVHTSGGLKLSFSCSNVNVNPFQCNPSPLQISSS <td>191</td> | 191 |
| QY <td>197<td>HAYEQLQE--GSMTTVSILLT-----TC-FQILYTLF<td>226</td></td></td> | 197 <td>HAYEQLQE--GSMTTVSILLT-----TC-FQILYTLF<td>226</td></td> | HAYEQLQE--GSMTTVSILLT-----TC-FQILYTLF <td>226</td> | 226 |
| DB <td>192<td>TALNLELTFEIGSLPEVMSIITLISYFILLTVVIRISKERKRYAFSTCTSHLMAVSLE<td>251</td></td></td> | 192 <td>TALNLELTFEIGSLPEVMSIITLISYFILLTVVIRISKERKRYAFSTCTSHLMAVSLE<td>251</td></td> | TALNLELTFEIGSLPEVMSIITLISYFILLTVVIRISKERKRYAFSTCTSHLMAVSLE <td>251</td> | 251 |
| QY <td>227<td>GLGTFEYVVRGGINMCCIIHALONINGFPSPSLNLFHTYVDKKAQISLVSGLW-NK<td>285</td></td></td> | 227 <td>GLGTFEYVVRGGINMCCIIHALONINGFPSPSLNLFHTYVDKKAQISLVSGLW-NK<td>285</td></td> | GLGTFEYVVRGGINMCCIIHALONINGFPSPSLNLFHTYVDKKAQISLVSGLW-NK <td>285</td> | 285 |
| DB <td>252<td>HGTVEMTFEQPANN-----PSLDKDKTMSLFTYV---IPMLNPLIYSWRNK<td>295</td></td></td> | 252 <td>HGTVEMTFEQPANN-----PSLDKDKTMSLFTYV---IPMLNPLIYSWRNK<td>295</td></td> | HGTVEMTFEQPANN-----PSLDKDKTMSLFTYV---IPMLNPLIYSWRNK <td>295</td> | 295 |
| QY <td>286<td>CYFALLVGLISLKDTLIQTIVGT<td>308</td></td></td> | 286 <td>CYFALLVGLISLKDTLIQTIVGT<td>308</td></td> | CYFALLVGLISLKDTLIQTIVGT <td>308</td> | 308 |
| DB <td>296<td>-----EVKDALHRAIAT<td>307</td></td></td> | 296 <td>-----EVKDALHRAIAT<td>307</td></td> | -----EVKDALHRAIAT <td>307</td> | 307 |

| RESULT | 15 |
|------------|--|
| ATP6_ANOOU | STANDARD; PRt: 226 AA. |
| AC | P33507; |
| DT | 01-FEB-1994 (Rel. 28, Created) |
| DT | 01-FEB-1994 (Rel. 28, Last sequence update) |
| DT | 15-JUL-1998 (Rel. 36, Last annotation update) |
| DE | ATP SYNTHASE A CHAIN (EC 3.6.1.34) (PROTEIN 6). |
| GN | ATP6. |
| OS | Anopheles quadrimaculatus (Mosquito). |
| OG | Mitochondrion. |
| OC | Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; |
| OC | Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae; |
| CC | Anopheles. |
| OX | NCBI_TaxID=7166; |
| RN | [1] |
| RP | SEQUENCE FROM N.A. |
| RC | STRAIN=ORLANDO; |
| RX | MEDLINE=92190510; PubMed=2134168; |
| RA | Cockburn A.F., Mitchell S.E., Seayright J.A.; |
| RT | "Cloning of the mitochondrial genome of Anopheles quadrimaculatus"; |
| RL | Arch. Insect Biochem. Physiol. 14:31-36(1992). |
| CC | -I- FUNCTION: KEY COMPONENT OF THE PROTON CHANNEL. IT MAY PLAY A |
| CC | DIRECT ROLE IN THE TRANSLLOCATION OF PROTONS ACROSS THE MEMBRANE. |
| CC | -I- SUBUNIT: F-TYPE ATPASES HAVE 2 COMPONENTS, CF(1) - THE CATALYTIC |
| CC | CORE - AND CF(0) - THE MEMBRANE PROTON CHANNEL. CF(1) HAS FIVE |
| CC | SUBUNITS: ALPHA(3), BETA(3), GAMMA(1), DELTA(1), EPSILON(1). CF(0) |
| CC | HAS THREE MAIN SUBUNITS: A, B AND C. |
| CC | -I- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. |

CC -1- SIMILARITY: BELONGS TO THE ATPASE A CHAIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: L04272; AAA9354.1; -
CC InterPro: IPR000568; ATP_synth_A.
CC Pfam: PF00119; ATP-synth_A: 1.
CC PRINTS: PR00123; ATPASEA.
CC PROSITE: PS00449; ATPASE_A: 1.
CC Hydrogen Ion transport: CF(0);
CC Mitochondrion; Transmembrane.
KW Hydrogen Ion transport; CF(0);
SQ SEQUENCE 226 AA; 25286 MW; FA4989EDAEFF1571 CRC64;

Query Match 5 7%; Score 94; DB 1; Length 226;
Best Local Similarity 23.9%; Pred. NO. 0.99; Mismatches 90; Indels 44; Gaps 10;
Matches 53; Conservative 35;
QY 49 LISNLFVLPFLOSSTTSHISFK--DAFLGILIPGYAALPNPWO--FSQFVKDLTK 104
DB 1 MMTNLFSEV---FDPSTTILNLSNWLSTFLGILLIPSEFWLPPNRQVWNNILLTLHK 56
QY 105 CVAMLL--TYCGPVLDV--LYHLNPKSSILEDYHEFLNIMSFNRFIFAPITEETPT 161
DB 57 EFKTLGSPSGHNGSTLMFISLFSLI-----MFNFFGLFPY-----IFTS 96
QY 162 SMLTTYLNLIPHSQVQQLFMQPSLFFGLAHNAHAYEQLQESMTTYSILLTFCQIL 221
DB 97 TSHLTTLALAFPLMLSLMGW-----INHTQHMFALHLPQGTTPVLMPPWVCLETT 149
QY 222 YTLFGGLTKFVFTGNTLMCCITLHALCNIMGFPSPRLN 263
DB 150 SNVIRPGT---LAVRLTANM---IAGHLLLTLLGNTGPMPTN 185

Search completed: February 12, 2002, 09:32:25
Job time: 341 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 12, 2002, 11:09:22 : Search time 277.71 Seconds
(without alignments)
5634.002 Million cell updates/sec

Title: US-09-165-460A-1

Perfect score: 1825
Sequence: 1 aactaccctttttctatct.....aaataacagattaacatt 1825

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 930621 seqs, 428662619 residues

Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

N.GeneSeq_1101:*

| | |
|-----|--|
| 1: | /SIDS2/gcgdata/geneseq/geneseqn/NA1980.DAT:* |
| 2: | /SIDS2/gcgdata/geneseq/geneseqn/NA1981.DAT:* |
| 3: | /SIDS2/gcgdata/geneseq/geneseqn/NA1982.DAT:* |
| 4: | /SIDS2/gcgdata/geneseq/geneseqn/NA1983.DAT:* |
| 5: | /SIDS2/gcgdata/geneseq/geneseqn/NA1984.DAT:* |
| 6: | /SIDS2/gcgdata/geneseq/geneseqn/NA1985.DAT:* |
| 7: | /SIDS2/gcgdata/geneseq/geneseqn/NA1986.DAT:* |
| 8: | /SIDS2/gcgdata/geneseq/geneseqn/NA1987.DAT:* |
| 9: | /SIDS2/gcgdata/geneseq/geneseqn/NA1988.DAT:* |
| 10: | /SIDS2/gcgdata/geneseq/geneseqn/NA1989.DAT:* |
| 11: | /SIDS2/gcgdata/geneseq/geneseqn/NA1990.DAT:* |
| 12: | /SIDS2/gcgdata/geneseq/geneseqn/NA1991.DAT:* |
| 13: | /SIDS2/gcgdata/geneseq/geneseqn/NA1992.DAT:* |
| 14: | /SIDS2/gcgdata/geneseq/geneseqn/NA1993.DAT:* |
| 15: | /SIDS2/gcgdata/geneseq/geneseqn/NA1994.DAT:* |
| 16: | /SIDS2/gcgdata/geneseq/geneseqn/NA1995.DAT:* |
| 17: | /SIDS2/gcgdata/geneseq/geneseqn/NA1996.DAT:* |
| 18: | /SIDS2/gcgdata/geneseq/geneseqn/NA1997.DAT:* |
| 19: | /SIDS2/gcgdata/geneseq/geneseqn/NA1998.DAT:* |
| 20: | /SIDS2/gcgdata/geneseq/geneseqn/NA1999.DAT:* |
| 21: | /SIDS2/gcgdata/geneseq/geneseqn/NA2000.DAT:* |
| 22: | /SIDS2/gcgdata/geneseq/geneseqn/NA2001.DAT:* |

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB | ID | Description |
|------------|-------|-------------|--------|----|----------|--------------------|
| 1 | 1802 | 98.7 | 1825 | 19 | AAV17601 | Yeast Afcl gene en |
| 2 | 140.4 | 7.7 | 1616 | 21 | AA36991 | Arabidopsis thalia |
| 3 | 137 | 7.5 | 936 | 22 | AA58252 | Oligonucleotide D1 |
| 4 | 137 | 7.5 | 936 | 22 | AA58254 | Oligonucleotide D1 |
| 5 | 137 | 7.5 | 936 | 22 | AA58257 | Oligonucleotide D1 |
| 6 | 137 | 7.5 | 936 | 22 | AA58259 | Oligonucleotide D2 |
| 7 | 137 | 7.5 | 936 | 22 | AA58262 | Oligonucleotide D2 |
| 8 | 137 | 7.5 | 936 | 22 | AA58265 | Oligonucleotide D1 |
| 9 | 136.8 | 7.5 | 936 | 22 | AA58252 | Oligonucleotide D1 |
| 10 | 136.8 | 7.5 | 936 | 22 | AA58254 | Oligonucleotide D1 |
| 11 | 136.8 | 7.5 | 936 | 22 | AA58257 | Oligonucleotide D1 |

| | | | | | | |
|----|-------|-----|-------|----|-----------|--------------------|
| 12 | 136.8 | 7.5 | 936 | 22 | AA58259 | Oligonucleotide D2 |
| 13 | 136.8 | 7.5 | 936 | 22 | AA58262 | Oligonucleotide D2 |
| 14 | 136.8 | 7.5 | 936 | 22 | AA58255 | Oligonucleotide D1 |
| 15 | 117.4 | 6.4 | 2966 | 20 | AA83493 | Human farnesylated |
| 16 | 117.4 | 6.4 | 2966 | 20 | AAV83752 | CDNA encoding huma |
| 17 | 106 | 5.8 | 2570 | 21 | AAFI8173 | Lung cancer associ |
| 18 | 88.2 | 4.8 | 684 | 21 | AAFI5007 | Trichoderma reesei |
| 19 | 79.8 | 4.4 | 649 | 21 | AAFI07931 | Fusarium venenatum |
| 20 | 72.2 | 4.0 | 844 | 22 | AAH08445 | Human CDNA clone (|
| 21 | 72.2 | 4.0 | 2161 | 22 | AAH18266 | Human CDNA sequenc |
| 22 | 66.8 | 3.7 | 1072 | 22 | AAE25691 | H. pylori HPC187 e |
| 23 | 63.6 | 3.5 | 1245 | 22 | AAE25625 | H. pylori HPS187 e |
| 24 | 61.6 | 3.4 | 1224 | 18 | AAE67867 | H. pylori cell env |
| 25 | 56.6 | 3.1 | 244 | 22 | AAE58238 | Oligonucleotide D1 |
| 26 | 56.2 | 3.1 | 310 | 22 | AAH71505 | Human cervical can |
| 27 | 53.8 | 2.9 | 464 | 22 | AAH33241 | Human colon cancer |
| 28 | 53 | 2.9 | 7204 | 20 | AAE23527 | O. longistamanta X |
| 29 | 52.8 | 2.9 | 1673 | 20 | AAE73320 | Human secreted pro |
| 30 | 52.2 | 2.9 | 244 | 22 | AAE58238 | Oligonucleotide D1 |
| 31 | 51.8 | 2.8 | 12020 | 21 | AAE83328 | ARSDRI polynucleot |
| 32 | 51.6 | 2.8 | 1641 | 17 | AAE15758 | Human interleukin- |
| 33 | 51.4 | 2.8 | 382 | 15 | AAQ55246 | Grapevine ribosome |
| 34 | 51.4 | 2.8 | 905 | 21 | AAE64642 | Partial sequence M |
| 35 | 50.8 | 2.8 | 612 | 22 | AAH71471 | Human cervical can |
| 36 | 50 | 2.7 | 797 | 22 | AAE02414 | Human secreted pro |
| 37 | 49.8 | 2.7 | 788 | 22 | AAH35003 | Human colon cancer |
| 38 | 49.6 | 2.7 | 1024 | 21 | AAE97189 | Human prostate can |
| 39 | 49.6 | 2.7 | 1459 | 22 | AAE02119 | Maize RAD51 orthol |
| 40 | 49.2 | 2.7 | 568 | 22 | AAI19154 | Probe #9087 for ge |
| 41 | 49.2 | 2.7 | 568 | 22 | AAI44305 | Human cervical can |
| 42 | 49.2 | 2.7 | 621 | 22 | AAH71551 | Human polynucleoti |
| 43 | 49.2 | 2.7 | 2040 | 22 | AAE63879 | Human secreted pro |
| 44 | 49 | 2.7 | 887 | 21 | AAE59257 | Human colon cancer |
| 45 | 48.8 | 2.7 | 208 | 21 | AAE98557 | |

ALIGNMENTS

| | | | | | | |
|----------|---|--|--|--|--|--|
| RESULT 1 | | | | | | |
| ID | AAV17601 | standard; DNA: 1825 BP. | | | | |
| XX | XX | | | | | |
| AC | AAV17601: | | | | | |
| XX | XX | | | | | |
| DT | 20-JUL-1998 | (first entry) | | | | |
| XX | XX | | | | | |
| DE | Yeast Afcl gene encoding Afclp protein. | | | | | |
| XX | XX | | | | | |
| KW | AFcl gene; Afclp; a-factor convertase; CNA protease; | | | | | |
| KW | zinc metalloprotease; yeast; prenylation; Ras; inhibitor; cancer; | | | | | |
| KW | colorectal carcinoma; pancreas carcinoma; leukemia; therapy; ss. | | | | | |
| OS | Saccharomyces cerevisiae. | | | | | |
| XX | XX | | | | | |
| FH | Key | Location/Qualifiers | | | | |
| FT | CDS | 342..1704 | | | | |
| FT | FT | /*lag= a | | | | |
| FT | FT | /transl_except= (pos:378..380, aa:Thr) | | | | |
| FT | FT | /transl_except= (pos:702..705, aa:Ser) | | | | |
| FT | FT | /note= "this codon has an apparent 1 nucleotide insertion, which alters the reading frame" | | | | |
| FT | FT | /transl_except= (pos:1255..1257, aa:Met) | | | | |
| FT | FT | /transl_except= (pos:1663..1665, aa:Leu) | | | | |
| PN | W09805786-A2. | | | | | |
| XX | XX | | | | | |
| PD | 12-FEB-1998. | | | | | |
| XX | XX | | | | | |
| PF | 06-AUG-1997: | 97WO-US14777. | | | | |
| XX | XX | | | | | |
| PR | 30-JUL-1997: | 97US-0902774. | | | | |

PR 07-AUG-1996; 96US-0023491.
 XX (REGC) UNIV CALIFORNIA.
 PA
 XX Ashby MN, Boyarchuk VL, Rine JD;
 PI
 XX WPI: 1998-145621/13.
 DR P-PSDB; AAM48301.
 XX
 PT Vector comprising nucleic acid coding for Afcip and Rceip proteins
 PT involved in protein prenylation - useful for screening for
 PT inhibitors of proteins, especially of mutated ras proteins involved
 PT in cancerous conditions
 XX
 PS
 XX
 CC Claim 11; Page 47-48; 59pp; English.
 CC
 CC This DNA sequence comprises the yeast Afc1 (a-factor convertase)
 CC gene that codes for a novel protein (see AAM48301), designated Afc1p,
 CC that mediates the removal of AAX tripeptide from CAAX proteins
 CC following prenylation. Afc1 was isolated using an autoocrine
 CC arrest, sensitised selection for CAAX proteases. This involved
 CC ecotopic expression of an a-factor in an a-cell and a CAAX
 CC permutational defective for proteolysis. Of 127 mutants isolated, 22
 CC had mutations in a single gene (AFC1). This is the first CAAX
 CC protease to be identified. Genetic knockout of Afc1 resulted in
 CC incomplete reduction of a-factor production, thereby exposing the
 CC presence of a second CAAX protease. The gene for this (see AAI17602)
 CC was named RCE1 (Ras and a-factor converting enzyme). Vectors
 CC comprising the Afc1 or Rce1 gene, polypeptides encoded by such
 CC vectors and recombinant cells transduced with the vectors are
 CC claimed. The novel Afc1p and Rce1p (see AAM48302) proteins can be
 CC used to screen for inhibitors of prenylation which can be used
 CC in the treatment of Ras dependent cancers such as colorectal and
 CC exocrine pancreatic carcinomas or myeloid leukaemia.
 XX
 SQ Sequence 1825 BP; 570 A; 398 C; 280 G; 577 T; 0 other;
 Query Match 98.7%; Score 1802; DB 19; Length 1825;
 Best Local Similarity 99.9%; Pred. No. 0;
 Matches 1824; Conservative 0; Mismatches 0; Indels 2; Gaps 2;

QY 481 ccactgtgctggaagacgaatgatgatgaacttttcaataatcaaggacatctcc 540
 DB 480 ccaactgtgctggaagacgaatgatgatgaacttttcaataatcaaggacatctcc 539
 QY 541 cgggccaaggccaagtctccatcttcgtgagcgtctataactagcccaaaagctagt 600
 DB 540 cgggccaaggccaagtctccatcttcgtgagcgtctataactagcccaaaagctagt 599
 QY 601 ttatcaataatgacctcttccctaataatcggcagatgagccgtttcttattgata 660
 DB 600 ttatcaataatgacctcttccctaataatcggcagatgagccgtttcttattgata 659
 QY 661 gtcccgccagtcagattcatatgtgtccacgtgcccagag-ttatgtcttctgg 719
 DB 660 gtcccgccagtcagattcatatgtgtccacgtgcccagagtttatgtcttctgg 719
 QY 720 tctctatccagttgtgtacctgtgtgatttggcaactctcttactataagcaattgt 779
 DB 720 tctctatccagttgtgtacctgtgtgatttggcaactctcttactataagcaattgt 779
 QY 780 cctggaagaaaatttggtttccaataattgacggtccaactatgatacagatagt 839
 DB 780 cctggaagaaaatttggtttccaataattgacggtccaactatgatacagatagt 839
 QY 840 caagagtcgactgtgagtgatgctgtgtggcccaactcttaccgttcccttaagat 899
 DB 840 caagagtcgactgtgagtgatgctgtgtggcccaactcttaccgttcccttaagat 899
 QY 900 ctgtataaatccctactgatttcccttggtaactatggtctctgttctgtgtcca 959
 DB 900 ctgtataaatccctactgatttcccttggtaactatggtctctgttctgtgtcca 959
 QY 960 aacttaagcctgaaacatcattccaggtcttcatatgagccatggttaataagttcaccc 1019
 DB 960 aacttaagcctgaaacatcattccaggtcttcatatgagccatggttaataagttcaccc 1019
 QY 1020 atggaagagcgtgagtcgaaagaaatcattcgaagtttggccgaagaagttggtccc 1079
 DB 1020 atggaagagcgtgagtcgaaagaaatcattcgaagtttggccgaagaagttggtccc 1079
 QY 1080 tctagataagaatttgttcatttgcagcgtcctcaaaaagatctctcatcaaacgcatatt 1139
 DB 1080 tctagataagaatttgttcatttgcagcgtcctcaaaaagatctctcatcaaacgcatatt 1139
 QY 1140 cagaggttgcatttccactcccaagagaaattgttgttgcacatttgaagacagtaa 1199
 DB 1140 cagaggttgcatttccactcccaagagaaattgttgttgcacatttgaagacagtaa 1199
 QY 1200 ttctactatgataaattcagcgtgttggcccaagaaatcgttgcacgtgcaaaaacaa 1259
 DB 1200 ttctactatgataaattcagcgtgttggcccaagaaatcgttgcacgtgcaaaaacaa 1259
 QY 1260 catcgttaataatgttcatctttagtcaatttgcacacttccatcttcccttccac 1319
 DB 1260 catcgttaataatgttcatctttagtcaatttgcacacttccatcttcccttccac 1319
 QY 1320 cagcatctacagaataatcatattttacaacacttggcgttttcttaagagaagttcac 1379
 DB 1320 cagcatctacagaataatcatattttacaacacttggcgttttcttaagagaagttcac 1379
 QY 1380 tggcagtttgttgaatcccggtatcaactaagaagatccccaattcatgtgattatgtt 1439
 DB 1380 tggcagtttgttgaatcccggtatcaactaagaagatccccaattcatgtgattatgtt 1439
 QY 1440 atttaagactatttaactccacacgcgaatgagcagatgagcagatgagttcaatttc 1499
 DB 1440 atttaagactatttaactccacacgcgaatgagcagatgagcagatgagttcaatttc 1499
 QY 1500 cagaactcatgaatatacaagctgattgtcttaaaaaaatctgggtacaaagcaaatct 1559
 DB 1500 cagaactcatgaatatacaagctgattgtcttaaaaaaatctgggtacaaagcaaatct 1559

PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148177.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149923.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158028.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.

PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 7.7%; Score 140.4; DB 21; Length 1616;
Best Local Similarity 48.4%; Pred.No.1,3e-23;
Matches 466; Conservative 0; Mismatches 451; Indels 45; Gaps 1;

QY 700 cagaattatgctcttgggtctcttaccagttgtctactgttgattgccactc 759
DB 422 catactcttctcttctgtgtgtgtatgatactgtccagatcactgttgcatt 481
QY 760 tcttactagcatttcttcttgaagaanaatttggtttcaataatttgcctgcca 819
DB 482 tcttctgactcaacttctgtctgagtcgagtcggcagtggttcaacaaacaaatag 541
QY 820 ctatgataccagataatgatacagaagctgacttggcgtatgcatgttggcccaatc 879
DB 542 atgttcattaggagatgatgatacagaacattccctctctgtcatatctagccaccatt 601
QY 880 cttaccctgttcccttaagaatcttgaataatccctactgatttcccttgttacctag 939
DB 602 gtgtgtcgataatttcatagtcagagaaggttcttacttctgtcactatctgttg 661
QY 940 gttctctgttctgttccaaatctttagcactgacaatccactgcttcttacc 999
DB 662 gaattcaagttatctcttctctctagtgatgatactataccctgttcttagaccg 721
QY 1000 atgttataagttacttccatcttggagagcggtgactgaaanaatctatgaagtgtg 1059
DB 722 ctcttcaacaagttacttccctcttccagatgtaggcctccgggagaagatgagaactt 781
QY 1060 gccgaataagttgggtccctctctatagataagatttttgcatttgcagctcaaaagactc 1119
DB 782 gcttctctctaaagtttcccttcttgaagaagctgttgttctgtatgatactcaagtgca 841
QY 1120 tctcatcaaacgcatcttccagagtttgcacttccactcccaagaagatttgttctg 1179
DB 842 agccataagaatgcttactatgatagttcttcttgaagaacaaagattgttcttattgat 901
QY 1180 gacactttagtgaacagtaattctactgataatgaatgaagcgttgttggccatgaatc 1239
DB 902 acgttgattcgcagtgcaagaatgagatgaatatttgggggtattgtgcaacagactt 961
QY 1240 ggtcactgycaaaanaaccacatgcttataatgttcatctttagtcaattgacaccttc 1299
DB 962 ggcacttgaacactgaatcacactacatactcgttcatgttcagttcaaatccctgtcttc 1021
QY 1300 ctcatcttcccttcttaccagcatctcaagaaatataatctttaaacaaccttggc 1359
DB 1022 ttacaatttgaagatacaacttctgcagaacatccatgatacttcttaagggttctgga 1081
QY 1360 ttcttctagaagaatccacactgycagtttctgttgcctgtatcatcaagaattcccc 1419
DB 1082 ttgatatacag-----cct 1096
QY 1420 attatcattgattatgtatttaacgacttataactcactcactgaaatgtgcaatgcaa 1479
DB 1097 gtctcactgttgtgatcataatttcaagcactgtgaatacactcactcactagtaagc 1156
QY 1480 ttgagatgagtttaatttccagactcctgaataatcagagttgttgcgttaagtaaaaa 1539
DB 1157 ttggccggaacctcgcttagtcgagcggtttagtcttggcgtgagtgcttggctgagag 1216
QY 1540 ttgggtctacaagcaaatctatgttagcgctcctaattgtactacaataacaaaccccttc 1599
DB 1217 ctgtgctatgcaaaaagactctgtccctgtctctagtgaacactcaggaagagaactatca 1276
QY 1600 accatgaatgtagctctgtatcttctagctataatattccatccaactcactagctgaa 1659
DB 1277 gcaatgaacactgtatcttactcagcttatacactacactacatccctcctctgttgaa 1336


```
QY 1383 cagtttgcacccgtatcaactaagaattcccatatcatgtattat 1442
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 440 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 381
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1443 taacgaacttaactcactcgaatgtgcataagcattcgtatgaatttccag 1502
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 380 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 321
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1503 aaccatgaataatcaagctgtatgtctataaaattgggtcacaagcaaatctatg 1562
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 320 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 261
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1563 taggctcctaattgatacaatacaaaaccttccacatgaatgtagctcctgta 1622
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 260 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 201
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1623 ttctagatacataatcccatcgaactctagctgaagaatcgaccgtctagactatgt 1682
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 200 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 141
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1683 tagtgaagaagaagaataatctatagatgatacatatagatgtagcgttaattc 1742
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 140 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 81
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1743 agcttcgtatcgtatatactacatacaacaggtatctactataagaataaaggaa 1802
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 80 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 21
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1803 gaaaaataaagaattaa 1821
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 20 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 2
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 6
AAF58259/c
ID AAF58259 standard; DNA; 936 BP.
XX
AC AAF58259;
XX
DT 24-APR-2001 (first entry)
XX
DE Oligonucleotide D2004.
XX
KM Electon-transfer group; ETM: mismatch; genotyping;
KM gene expression; ss.
XX
OS Synthetic.
XX
PN WO200107665-A2.
XX
PD 01-FEB-2001.
XX
PF 26-JUL-2000; 2000WO-US20476.
XX
PR 26-JUL-1999; 99US-0145695.
PR 17-MAR-2000; 2000US-0190259.
XX
PA (CLIN-) CLINICAL MICRO SENSORS INC.
XX
PI Umek RM;
XX
DR WPI: 2001-159728/16.
XX
PT Nucleic acids containing electron-transfer group, useful as labels in
PT hybridization assays, e.g. for genotyping, allowing repeat analyses on
PT a single surface
XX
PS Example 6; Page 128; 159pp; English.
XX
CC The present invention relates to a composition comprising two nucleic
CC acids each containing an electron-transfer group (ETM) having
CC different redox potentials. The invention is used for electronic
```

```
CC detection of nucleic acids, especially of substitutions (mismatches)
CC and single-nucleotide polymorphisms, e.g. for genotyping.
CC monitoring gene expression.
XX
SO Sequence 936 BP; 6 A; 138 C; 8 G; 8 T; 776 other;

Query Match 7.5%; Score 137; DB 22; Length 936;
Best Local Similarity 1.4%; Pred. No. 6.6e-23;
Matches 11; Conservative 499; Mismatches 289; Indels 0; Gaps 0;

QY 1023 ggaggcgttgaactgaaataatctatgaagtgtggccatagatgtgggtccctc 1082
      || || | : : : : : : : : : : : : : : : : : : : : : :
DB 800 GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 741
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1083 agataagattttgtcatatgacgctcaaaaagatcttcattcaacgcataattcac 1142
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 740 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 681
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1143 aggttgcattcaacctcagaagatgtttgttgacactttagtgaacagtaattc 1202
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 660 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 621
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1203 tactgatgaattacgctgttttggccatgaatcgctcactgcaaaaaccacat 1262
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 620 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 561
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1263 cgttaattgctcatcttagtcaattgcacaccttcctcatcttccctttccag 1322
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 560 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 501
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1323 catctacagaatacatatttacaacaccttggtcttcttagagaagttccactg 1382
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 500 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 441
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1383 cagtttgcacccgtatcaactaagaattcccatatcatgtattat 1442
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 440 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 381
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1443 taacgaacttaactcactcgaatgtgcataagcattcgtatgaatttccag 1502
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 380 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 321
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1503 aactatgaataatcaagctgtatgtctataaaattgggtcacaagcaaatctatg 1562
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 320 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 261
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1563 taggctcctaattgatacaatacaaaaccttccacatgaatgtagctcctgta 1622
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 260 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 201
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1623 ttctagatacataatcccatcgaactctagctgaagaatcgaccgtctagactatgt 1682
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 200 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 141
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1683 tagtgaagaagaagaataatctatagatgatacatatagatgtagcgttaattc 1742
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 140 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 81
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1743 agcttcgtatcgtatatactacatacaacaggtatctactataagaataaaggaa 1802
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 80 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 21
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1803 gaaaaataaagaattaa 1821
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 20 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 2
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 7
AAF58262/c
ID AAF58262 standard; DNA; 936 BP.
XX
AC AAF58262;
```

```

XX 24-APR-2001 (first entry)
XX 24-APR-2001 (first entry)
XX Oligonucleotide D2007.
XX Electron-transfer group: ETM; mismatch; genotyping;
XX gene expression; ss.
XX Synthetic.
XX WO200107665-A2.
XX 01-FEB-2001.
XX 26-JUL-2000; 2000WO-US20476.
XX 26-JUL-1999; 99US-0145695.
XX 17-MAR-2000; 2000US-0190259.
XX (CLIN-) CLINICAL MICRO SENSORS INC.
XX Umek RM;
XX WPI: 2001-159728/16.
XX Nucleic acids containing electron-transfer group, useful as labels in
XX hybridization assays, e.g. for genotyping, allowing repeat analyses on
XX a single surface.
XX Example 6; Page 128; 159pp; English.
XX The present invention relates to a composition comprising two nucleic
XX acids each containing an electron-transfer group (ETM) having
XX different redox potentials. The invention is used for electronic
XX detection of nucleic acids, especially of substitutions (mismatches)
XX and single-nucleotide polymorphisms, e.g. for genotyping,
XX monitoring gene expression.
XX Sequence 936 BP; 5 A; 139 C; 10 G; 6 T; 776 other;

```

```

Query Match 7.5%; Score 137; DB 22; Length 936;
Best Local Similarity 1.4%; Pred. No. 6.6e-23;
Matches 11; Conservative 499; Mismatches 289; Indels 0; Gaps 0;
OY 1023 gagagacgtgacacgaaataatcatgaaagttgacgagatggtggtccctc 1082
DB 800 GCGGGGGGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 741
OY 1083 agataagattttgtcatgacgctcaaaagatctctcattcaaacgacattcac 1142
DB 740 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 681
OY 1143 aggttgcattaccctccagaagaatgtttgttcgacattagtagaagaatttc 1202
DB 680 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 621
OY 1203 tactgatgaattacgctgttttggcccatgaaatcgctcactgacaaataaccacat 1262
DB 620 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 561
OY 1263 cgtataatgcatctttagtcaattgacacactctcctcattctcccttaccag 1322
DB 560 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 501
OY 1323 catctacagaatacatcatttacaacacttcgcttctctagagaagttccactg 1382
DB 500 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 441
OY 1383 cagtttctgacccggtatcaccaagaagaattcccatcatcattgattatgttatt 1442
DB 440 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 381

```

```

OY 1443 taagacttaactccacgtcgaatggtccatcgatgagtttaattccag 1502
DB 380 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 321
OY 1503 aactatgaataatcagcgtgactatgcttaaaaaattgggtaccagaacatctag 1562
DB 320 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 261
OY 1563 tagggtcttaattgcttacaacaaacacttccacatgaaatgtagatcctctgta 1622
DB 260 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 201
OY 1623 ttctagctatcatttcccatccacactagctgaagaatcgacgtcttagactagt 1682
DB 200 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 141
OY 1683 tagtgaagaagaagaactaactatcagagtagacacatattagcattgaatttc 1742
DB 140 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 81
OY 1743 agcttcgttatgtcatatcatcacatcacacaggtatctactataagaataagaaa 1802
DB 80 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 21
OY 1803 gaaaaataaacgattaaa 1821
DB 20 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 2

```

```

RESULT 8
AAF58255/C
ID AAF58255 standard; DNA; 938 BP.
AC AAF58255;
XX 24-APR-2001 (first entry)
XX Oligonucleotide D1876.
XX Electron-transfer group: ETM; mismatch; genotyping;
XX gene expression; ss.
XX Synthetic.
XX WO200107665-A2.
XX 01-FEB-2001.
XX 26-JUL-2000; 2000WO-US20476.
XX 26-JUL-1999; 99US-0145695.
XX 17-MAR-2000; 2000US-0190259.
XX (CLIN-) CLINICAL MICRO SENSORS INC.
XX Umek RM;
XX WPI: 2001-159728/16.
XX Nucleic acids containing electron-transfer group, useful as labels in
XX hybridization assays, e.g. for genotyping, allowing repeat analyses on
XX a single surface.
XX Example 6; Page 127; 159pp; English.
XX The present invention relates to a composition comprising two nucleic
XX acids each containing an electron-transfer group (ETM) having
XX different redox potentials. The invention is used for electronic
XX detection of nucleic acids, especially of substitutions (mismatches)
XX and single-nucleotide polymorphisms, e.g. for genotyping,
XX monitoring gene expression.
XX Sequence 938 BP; 4 A; 144 C; 9 G; 5 T; 776 other;

```

| | |
|----|--|
| XX | Electron-transfer group; ETW; mismatch; genotyping; |
| KM | gene expression; ss. |
| XX | |
| OS | Synthetic. |
| XX | |
| PN | WO200107665-A2. |
| XX | |
| PD | 01-FEB-2001. |
| XX | |
| PE | 26-JUL-2000; 2000WO-US20476. |
| XX | |
| PR | 26-JUL-1999; 99US-0145695. |
| PR | 17-MAR-2000; 2000US-0190259. |
| XX | |
| PA | (CLIN-) CLINICAL MICRO SENSORS INC. |
| XX | |
| PI | Unex RM; |
| PT | |
| DR | WPI; 2001-159728/16. |
| XX | |
| PT | Nucleic acids containing electron-transfer group, useful as labels in |
| PT | hybridization assays, e.g. for genotyping, allowing repeat analyses on |
| PT | a single surface - |
| XX | |
| PS | Example 6; Page 127; 159pp; English. |
| XX | |
| CC | The present invention relates to a composition comprising two nucleic |
| CC | acids each containing an electron-transfer group (ETW) having |
| CC | different redox potentials. The invention is used for electronic |
| CC | detection of nucleic acids, especially of substitutions (mismatches) |
| CC | and single-nucleotide polymorphisms, e.g. for genotyping, |
| CC | monitoring gene expression. |
| XX | |
| XQ | Sequence 936 BP; 4 A; 139 C; 10 G; 7 T; 776 other; |

SQ Sequence 936 BP; 4 A; 139 C; 10 G; 7 T; 776 other;

| | | | | |
|-----------------------|------|------------------|----------------|------------|
| Query Match | 7.5% | Score 136.8 | DB 22 | Length 936 |
| Best Local Similarity | 0.8% | Pred No. 7.4e-23 | | |
| Matches | 6 | Conservative 496 | Mismatches 278 | Indels 0 |
| | | | | Gaps 0 |

| | | | |
|----|------|---|------|
| QY | 1034 | aactgaaaaaatcttgaaggtttgcgataagatcttggtgtccctctagaataagttt | 1093 |
| Db | 8 | www..... | 67 |
| QY | 1094 | ttgtcatctgacgctcaaaaagactctctcatcaaacgcatattccagtttgcgat | 1153 |
| Db | 68 | www..... | 127 |
| QY | 1154 | tcacctccagaagaattgtttgtctgacactttgtacagtaattctactgtatgaa | 1213 |
| Db | 128 | www..... | 187 |
| QY | 1214 | ttacgctgttttggcccatgaaatcgtgcactgagcaaaaaacacatcgttaatactg | 1273 |
| Db | 188 | www..... | 247 |
| QY | 1274 | tcactttgttcgaatggaacacttccatcttctccctttcacccgacatctcaaa | 1333 |
| Db | 248 | www..... | 307 |
| QY | 1334 | atacatcaatttcaaacacctgcgctttctctagagaagccacgtgcagtttctgt | 1393 |
| Db | 308 | www..... | 367 |
| QY | 1394 | atcccgltacactaaagaaatcccatcatcatcattggaatttgtatttaagactat | 1453 |
| Db | 368 | www..... | 427 |
| QY | 1454 | taactcaactcgaattgtgcatacgaatctcgttgatgatttaattccagaactatgat | 1513 |
| Db | 428 | www..... | 487 |
| QY | 1514 | atcaagctgatcttatctgtaaaaaatctgggctacagaacaaatctatgtatgggcttaa | 1573 |

```
Db 488 www. .... 547
Oy 1574 ttgattcacaataaaccttccaccatgagttagctctgattctgatac 1633
Db 548 www. .... 607
Oy 1634 attatcccatccactagctgaagaatcgacgcgtctgactagttagtgaaga 1693
Db 608 www. .... 667
Oy 1694 agaaaactaactatagtagacacatatagcgttgaatcagcttgat 1753
Db 668 www. .... 727
Oy 1754 gtctatactacatacaccaggtatctactataagaagaagaataaa 1813
Db 728 www. .... 787

RESULT 10
AAF58254
ID AAF58254 standard; DNA; 936 BP.
XX
AC AAF58254;
XX
DT 24-APR-2001 (first entry)
XX
DE Oligonucleotide D1875.
XX
KW Electron-transfer group; ETM; mismatch; genotyping;
KM gene expression; ss.
XX
OS Synthetic.
XX
PN WO200107665-A2.
XX
PD 01-FEB-2001.
XX
PE 26-JUL-2000; 2000MO-US20476.
XX
PR 26-JUL-1999; 99US-0145695.
XX
PR 17-MAR-2000; 2000US-0190259.
XX
PA (CLIN-) CLINICAL MICRO SENSORS INC.
XX
PI Umek RM;
XX
DR WPI; 2001-159728/16.
XX
PT Nucleic acids containing electron-transfer group, useful as labels in
PT hybridization assays, e.g. for genotyping, allowing repeat analyses on
PT a single surface
XX
PS Example 6; Page 127; 159pp; English.
XX
CC The present invention relates to a composition comprising two nucleic
CC acids each containing an electron-transfer group (ETM) having
CC different redox potentials. The invention is used for electronic
CC detection of nucleic acids, especially of substitutions (mismatches)
CC and single-nucleotide polymorphisms, e.g. for genotyping,
CC monitoring gene expression.
XX
SQ Sequence 936 BP; 4 A; 144 C; 7 G; 5 T; 776 other;
```

```
Query Match 7.5%; Score 136.8; DB 22; Length 936;
Best Local Similarity 0.8%; Pred. No. 7.4e-23;
Matches 6; Conservative 496; Mismatches 278; Indels 0; Gaps 0;
Oy 1034 aactgaataaactatctgaaagtgtgcccagataggtggttccctcagataaagatt 1093
Db 8 www. .... 67
```

```
Oy 1094 ttgattcagcgtcacaagaatctctcatctcaaacgcataattcagagttgcat 1153
Db 68 www. .... 127
Oy 1154 tcacctcaagaagaattgttcttcgaccttagtgaacagtaattctactatgaa 1213
Db 128 www. .... 187
Oy 1214 ttacggtgttttggccatgaatcggtcactgccaataaacacatcgttaabtg 1273
Db 188 www. .... 247
Oy 1274 tcaatttagtcaattgcacaccttcctcattcttccttccattcaccagcctacagaa 1333
Db 248 www. .... 307
Oy 1334 atacatcatttacaacaccttcgcttctctcagaagaagtcacatgacgttcttg 1393
Db 308 www. .... 367
Oy 1394 atccggtatacctaagaatcccatlaltcattgattatgatttaaacactat 1453
Db 368 www. .... 427
Oy 1454 taactcactcgaatggtccatgcattcgatgagtttaattccaagaatcagat 1513
Db 428 www. .... 487
Oy 1514 atcaagctgattcttataaataatggtctacaagaacatctatgtaggctcaa 1573
Db 488 www. .... 547
Oy 1574 ttgattcaaaatcaaaaccttcacacatgaatgtagctctgtatctcagctac 1633
Db 548 www. .... 607
Oy 1634 attatcccatccactagctgaagaatcgacgcgtctgactagttagtgaaga 1693
Db 608 www. .... 667
Oy 1694 agaaaactaactatagtagacacatatagcgttgaatcagcttgat 1753
Db 668 www. .... 727
Oy 1754 gtctatactacatacaccaggtatctactataagaagaagaataaa 1813
Db 728 www. .... 787

RESULT 11
AAF58257
ID AAF58257 standard; DNA; 936 BP.
XX
AC AAF58257;
XX
DT 24-APR-2001 (first entry)
XX
DE Oligonucleotide D1954.
XX
KW Electron-transfer group; ETM; mismatch; genotyping;
KM gene expression; ss.
XX
OS Synthetic.
XX
PN WO200107665-A2.
XX
PD 01-FEB-2001.
XX
PR 26-JUL-2000; 2000MO-US20476.
XX
PR 26-JUL-1999; 99US-0145695.
XX
PR 17-MAR-2000; 2000US-0190259.
```


1694 agaaaaaactaatctatagagtacacacataatgcatgtacgcgttaattcagcttcgttat 1755

CC protein (AA26897) can be used to design specific inhibitors of enzymes
CC involved in degradation of prenylated proteins, particularly Ras and
CC related proteins and for treating disease mediated by Ras and related
CC proteins and/or the genes encoding them, specifically for diagnosis and
CC treatment of cancers.

XX Sequence 2966 BP; 844 A; 515 C; 578 G; 1029 T; 0 other;

Query Match 6.4%; Score 117.4; DB 20; Length 2966;
Best Local Similarity 50.5%; Pred. No. 4.1e-18;
Matches 370; Conservative 0; Mismatches 351; Indels 12; Gaps 3;

```
QY 471 aacaagtgcacccgtgctggaagacgaacatgataagaaacttgcataaacaag 530
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 177 aactcatgtaccaccggaagttagacagatcatgatactctgaacattgagaactcg 236
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 531 gaactactccggcgaagcccaagttctcatttcggtgacgtctataacctagcca 590
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 237 actctacaactgataaagaacattcagcttcgtctcaggaactctatcagagactga 296
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 591 aagctagtttcatcaataagacactctccctaataatcgcacatgacggttctt 650
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 297 aggcactcttactctctcttctggaagataactctatctctggaactcttgacggtt 356
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 651 attgaatgcagtcctgcacagatcatcaattcagtcctcactgacgacagattatg 710
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 357 ct--gtgttatgctggtcttggaacgaataatgatacactcagctccctggttctgt 414
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 711 ctcttggtctctatccagtttctcacttggtgattgacactctctactatag 770
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 415 ctgltggtctacact-----ttcagtcgactgactgttgcacatgagcttataa 467
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 771 ccatttgcctggagaagaataattggttcaataaattgacccgaactatgatacac 830
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 468 tactcttgagagaagaanaacatgcttcaatacaacagacttgggttcttcatgaa 527
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 831 cgatagatcaagaagatcgcgacttgcglatgataatggtggccaatccattacatt 890
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 528 agatgcaatcaagaattgtgtgacacagatgatttgtgctgtgtcttcaactct 587
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 891 cctaagatcttgaataatccactgatttcccttggtatcattatgctctctgt 950
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 588 acttacaattataaattgggggtgactatttttattattatgctgtgcttcaact 647
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 951 cgttgtccaatcttagccatgacataatccatgcttcatcagcccatgattataa 1010
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 648 agtctgctctggttcttgcaacaatctatgctgattatattgccccttattgacaa 707
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1011 gtccactcatttggaagcgttgaactgaanaaaatcattgaagaatttggccgataag 1070
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 708 attcacactctgctgtgggaagcgttaagaagaagaattgaagtaatgacgaagaat 767
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1071 tgggtccctctagataagaatttgtcatgacggtcgaanaagaatcttcatcaca 1130
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 768 tgacttcccttgacgaagtgatgttgaagaagactaaacgctctccacagcaa 827
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1131 cgcataattcaaggttgcacatcactcaagaagaattgttgtcgcacatttgt 1190
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 828 tgcctattcttaaggctt---ctcaagaacacagcaatgatttgttgcacacttact 884
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1191 gaacagtaattct 1203
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 885 agaagagtaactct 897
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

Search completed: February 12, 2002, 13:25:22
Job time: 8160 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 12, 2002, 10:13:07 ; Search time 2973.29 Seconds

(without alignments)
6595.739 Million cell updates/sec

Title: US-09-165-460A-1

Perfect score: 1825
Sequence: 1 acctaccctttttctatct.....aaataacgattaacatt 1825

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

EST:*
1: em_estfun:*
2: em_esthum:*
3: em_estin:*
4: em_estom:*
5: em_estpl:*
6: em_estba:*
7: em_estro:*
8: em_estov:*
9: em_hic:*
10: 9b_est1:*
11: 9b_est2:*
12: 9b_hic:*
13: 9b_gss:*
14: em_gss_fun:*
15: em_gss_hum:*
16: em_gss_hiv:*
17: em_gss_pln:*
18: em_gss_pro:*
19: em_gss_rtd:*
20: em_gss_vrt:*
21: em_gss_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|--------------------|
| 1 | 283.8 | 15.6 | 819 | 13 | ALA20625 T7 end of |
| 2 | 275.8 | 15.1 | 985 | 13 | ALA33329 T7 end of |
| 3 | 267 | 14.6 | 516 | 13 | AZ925942 476.d1005 |
| 4 | 259.8 | 14.2 | 328 | 13 | AZ923982 4906.1b97 |
| 5 | 238.8 | 13.1 | 893 | 13 | CNS06CIS |
| 6 | 213.6 | 11.7 | 517 | 13 | AZ930382 474.dhz53 |
| 7 | 177.2 | 9.7 | 1116 | 13 | AL393312 T7 end of |
| 8 | 161.6 | 8.9 | 1081 | 13 | CNS06CIS |
| 9 | 108.4 | 5.9 | 859 | 11 | BGI76721 T7 end of |
| 10 | 106.8 | 5.9 | 424 | 10 | AA965341 602313533 |
| 11 | 105.6 | 5.8 | 543 | 11 | BF051006 EST436181 |
| 12 | 102.8 | 5.6 | 708 | 10 | AM174887 fe06h01.y |

| | | | | | | |
|----|------|-----|------|----|----------|---------------------|
| 13 | 97 | 5.3 | 394 | 11 | T37813 | T37813 EST103073 S |
| 14 | 96.8 | 5.3 | 371 | 10 | AA787339 | AA787339 n3b02a1.r |
| 15 | 95.6 | 5.2 | 882 | 13 | CNS01F15 | ALA14258 Anopheles |
| 16 | 95.4 | 5.2 | 586 | 11 | AM146879 | AM146879 614089B06 |
| 17 | 92.4 | 5.1 | 651 | 11 | BG557737 | BG557737 EML_56-C1 |
| 18 | 92.2 | 5.1 | 491 | 10 | BE459882 | BE459882 EST415174 |
| 19 | 92 | 5.0 | 493 | 11 | BG556491 | BG556491 EML_37-G1 |
| 20 | 90.6 | 5.0 | 589 | 11 | AM585792 | AM585792 EST317415 |
| 21 | 87 | 4.8 | 430 | 10 | AM759861 | AM759861 s155a11.y |
| 22 | 84 | 4.6 | 531 | 10 | BE054474 | BE054474 GA_Ea001 |
| 23 | 83.4 | 4.6 | 627 | 11 | BE205165 | BE205165 EST397841 |
| 24 | 83.2 | 4.6 | 457 | 11 | BG651426 | BG651426 sad46e11. |
| 25 | 82.2 | 4.5 | 519 | 10 | AI328626 | AI328626 asd04ne.f |
| 26 | 81.2 | 4.4 | 388 | 10 | AM759000 | AM759000 s133d11.y |
| 27 | 81.2 | 4.4 | 748 | 11 | BG836625 | BG836625 Zm08_0390 |
| 28 | 80.6 | 4.4 | 472 | 10 | AI330202 | AI330202 c4d10ne.f |
| 29 | 80.6 | 4.4 | 600 | 10 | AI066958 | AI066958 HOR-10 C4 |
| 30 | 78.6 | 4.3 | 648 | 11 | BG874694 | BG874694 RC2-BN007 |
| 31 | 78 | 4.3 | 699 | 10 | AI759788 | AI759788 sbd4c03.y |
| 32 | 75.2 | 4.1 | 471 | 10 | AM678001 | AM678001 WSI_12-H0 |
| 33 | 75.2 | 4.1 | 499 | 10 | AM677942 | AM677942 WSI_12-H0 |
| 34 | 75.2 | 4.1 | 871 | 13 | AZ685824 | AZ685824 ENTH54TF |
| 35 | 73.6 | 4.0 | 706 | 10 | AI759796 | AI759796 sbd4d03.y |
| 36 | 73.4 | 4.0 | 600 | 10 | AM520069 | AM520069 660056F02 |
| 37 | 73.2 | 4.0 | 699 | 11 | BG862090 | BG862090 602795627 |
| 38 | 72.8 | 4.0 | 280 | 10 | AM120333 | AM120333 614089B06 |
| 39 | 72.8 | 4.0 | 440 | 10 | AI757461 | AI757461 ERTS2e05 |
| 40 | 72.4 | 4.0 | 1224 | 10 | BE035468 | BE035468 MO06D10 M |
| 41 | 71 | 3.9 | 446 | 10 | AM459275 | AM459275 sh22h03.y |
| 42 | 70.8 | 3.9 | 603 | 10 | BE060801 | BE060801 HVSMEG001 |
| 43 | 65.4 | 3.6 | 609 | 10 | AU037739 | AU037739 AU037739 |
| 44 | 65.4 | 3.6 | 805 | 13 | CNS009F8 | AL053500 Drosophila |
| 45 | 65 | 3.6 | 365 | 10 | AL515373 | AL515373 AL515373 |

ALIGNMENTS

| | | | | | |
|------------|--|-------------|-----|-----|----------------------|
| RESULT 1 | CNS06Y63 | 819 bp | DNA | GSS | 06-JUL-2001 |
| LOCUS | T7 end of clone AY00A006E12 of library AY00A | | | | from strain CBS 6340 |
| DEFINITION | of Kluyveromyces thermotolerans, genomic survey sequence. | | | | |
| ACCESSION | AL420625 | | | | |
| VERSION | AL420625.1 | GI:12203813 | | | |
| KEYWORDS | GSS. | | | | |
| SOURCE | Kluyveromyces thermotolerans. | | | | |
| ORGANISM | Kluyveromyces thermotolerans | | | | |
| REFERENCE | Eukaryota; Fungi; Ascomycota; Saccharomycetaceae; Kluyveromyces. | | | | |
| AUTHORS | Malpertuy,A., Llorente,B., Blandin,G., Artiguenave,F., Wincker,P. | | | | |
| TITLE | Genomic exploration of the hemiascomycetous yeasts: 10. | | | | |
| JOURNAL | Kluyveromyces thermotolerans | | | | |
| MEDLINE | FEBS Lett. 487 (1), 61-65 (2000) | | | | |
| AUTHORS | 2 (bases 1 to 819) | | | | |
| REFERENCE | Souciot,J.L., Aigle,M., Artiguenave,F., Blandin,G., de Montigny,J., Dujon,B., Durieux,P., Casaregola,S., Maupertuy,A., Neveuillise,C., Ozier-Kalogeropoulos,O., Potier,S., Saurin,W., Tekala,F., Toffano-Nicolas,C., Wesolowski-Louvel,M., Wincker,P. and Weissenbach,J. | | | | |
| JOURNAL | Genomic exploration of the hemiascomycetous yeasts: 1. A set of yeast species for molecular evolution studies | | | | |
| MEDLINE | FEBS Lett. 487 (1), 3-12 (2000) | | | | |
| AUTHORS | 20584720 | | | | |
| REFERENCE | 3 (bases 1 to 819) | | | | |
| JOURNAL | Genoscope. | | | | |
| MEDLINE | Submitted (08-SEP-2000) Genoscope - Centre National de Sequencage, | | | | |
| AUTHORS | Direct Submission | | | | |
| JOURNAL | Submitted (08-SEP-2000) Genoscope, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail : | | | | |

COMMENT

segrefgenoscope.cns.fr - Web : www.genoscope.cns.fr)
 This GSS is part of a random genomic sequencing program of thirteen yeast species: *Saccharomyces bayanus* var. *uvatum*, *Saccharomyces exiguus*, *Saccharomyces servazzii*, *Zygosaccharomyces rouxii*, *Saccharomyces kluyveri*, *Kluyveromyces thermotolerans*, *Kluyveromyces lactis* var. *lactis*, *Kluyveromyces marxianus* var. *marxianus*, *Pichia angusta*, *Debaryomyces hansenii* var. *hansenii*, *Pichia sorbitophila*, *Candida tropicalis* and *Yarrowia lipolytica*. Genomic inserts of 3 to 5 kb were prepared and both extremities were sequenced. See keywords for description of this sequence and for the sequence of the other extremity of this insert.

FEATURES

Source

Location/Qualifiers
 1..819
 /organism="Kluyveromyces thermotolerans"
 /strain="CBS 6340"
 /db_xref="taxon:4916"
 /clone="AY0A006E12"
 /clone_lib="AY0A"
 /note="end : 17"
 <20..>817

misc_feature

/note="similar to *Saccharomyces cerevisiae* ORF YJR117w [STE24 : zinc metallo-protease]"
 /evidence=not_experimental

BASE COUNT 187 a 219 c 195 g 209 t 9 others

ORIGIN

Query Match

Best Local Similarity 15.6% Score 283.8; DB 13; Length 819;
 Matches 484; Conservative 3; Mismatches 320; Indels 3; Gaps 1;

```

Oy 354 taagaagattctgcacatccataatcccgtagaataatcaatcttcggttcgatt 413
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 10 TYAAGAGAGATTGATCGCCCAATATGCGAATGCAATGATGATGATGATGATGAT 69
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Oy 414 tgcacaattcttcgaactcttgaagtagacagtagacagtagacagtagacagtag 473
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 70 GGGCCAGTTCTGCTCGACACGTCAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCT 129
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Oy 474 aaagtgcacactgtagaagtagaagtagaagtagaagtagaagtagaagtagaagtag 533
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 130 GCAGCTTCTTAAGTACTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 189
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Oy 534 ctactccgggccaagggccaaggtctccattcttcgtagcgtcctaactagcccaaaa 593
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 190 GTATTGCGAGAGCCAAAGCAATCTCTCATCTGATCTGATCTGATCTGATCTGATCT 249
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Oy 594 gctagtttcaataaataagacctctccctaataatctgtagacatgagccgtttcttatt 653
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 250 CCYTGTATTTTAAAGTGGATTTGTGGCCCGCATTTGGCAGATGGGTGAGGCGCTGCG 309
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Oy 654 gaatgacgttcgtagaagtagaagtagaagtagaagtagaagtagaagtagaagtagaagtag 713
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 310 GCA---GAAGCTTCGGCTAAAGATTGGCGGGGCTCTTACAAATTTGGCCAGATTGCTGTT 366
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Oy 714 ctggtgctctatccagattgtctacacttggttgattgtagcactctctactatagca 773
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 367 CTTGACGGNGCTCACCAACATCTCTCTTACACCTCCACACCTCTTACTACAGGA 426
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Oy 774 ttttctctgagaagaataattggtttcaataatgtagcgtcgaactatgtagacaga 833
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 427 CTTTGGCTTAGAAGAGAGTTGGCTTCACAAAGCTGAGCGTGAATTTGTGATACATGA 486
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Oy 834 tatgtcaagaagctgtagattgtagtctatgtagtcccaatcccttacttcttct 893
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 487 CAATAATCAAAAGGCACTATTCTGAGCGGCGGCGGCTTTGGCCCTTGATGCGCNCCTTT 546
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Oy 894 taagaattctgtaataatccctactgattctccttgtagacattatgtagtctctgct 953
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 547 GAAAGGCTTGCAGGCTTTCCCGACGACTCTTGTGGNACATCTCTCTTATCTGCG 606
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Oy 954 tttccaatcttagcattgacaatcatctccagttcttcatcattgcccatttaagt 1013
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

Db

607 ACTGCAAGTTCTACGAANTGCTGCTGATCTGCTACATTTATGCACTTTTCAACAGTT 666

Oy

1014 cactccattggaagcgtgtaacttgaataaacttgaagttgagccgtagagttgg 1073

Db

667 CACGGCATTGAGGAGGCGGAACCTACAGAGCTGATCGAGGCTTGAGCTTACCGGCTGG 726

Oy

1074 gtccctctagataagatttgcattatgtagcgtcgaacaaagattctcattcaaacgc 1133

Db

727 CTTCACACTGCATCAAAATTTGCTGTGTGAGCGGCTGCACAAAGCTTATCCCATCAAGC 786

Oy

1134 atattcaagattgtagcattcattcaacaa 1163

Db

787 CTACTTCACTGCTGCTGCTTCACTACAGCA 816

FEATURES

Source

Location/Qualifiers
 1..985
 /organism="Pichia angusta"
 /strain="CBS 4732"
 /db_xref="taxon:4905"
 /clone="BB0A015F09"
 /clone_lib="BB0A"
 /note="end : 17"
 <86..>973

misc_feature

/note="similar to *Saccharomyces cerevisiae* ORF YJR117w [STE24 : zinc metallo-protease]"
 /evidence=not_experimental

REFERENCE

1 (bases 1 to 985)
 Blandin,G., Llorente,B., Malpertuy,A., Wincker,P., Artiguenave,F. and Dujon,B.
 Genomic exploration of the hemiascomycetous yeasts: 13. *Pichia angusta*

TITLE

Genomic exploration of the hemiascomycetous yeasts: 13. *Pichia angusta*

JOURNAL

FEMS Lett. 487 (1), 76-81 (2000)

MEDLINE

20584723

REFERENCE

2 (bases 1 to 985)
 Soucieu,J.L., Aigle,M., Artiguenave,F., Blandin,G., Boiotin-Pukhara,M., Bon,E., Brothier,P., Casaregola,S., de-Montigny,J., Dujon,B., Durieux,P., Lepingle,A., Llorente,B., Malpertuy,A., Neugeglise,C., Ozier-Kalogeropoulos,O., Potier,S., Saurin,W., Tekala,F., Toffano-Nicohe,C., Wesolowski-Louvel,M., Wincker,P. and Weissendbach,J.
 Genomic exploration of the hemiascomycetous yeasts: 1. A set of yeast species for molecular evolution studies

JOURNAL

FEMS Lett. 487 (1), 3-12 (2000)

MEDLINE

20584711

REFERENCE

3 (bases 1 to 985)
 Genoscope.
 Direct Submission
 Submitted (08-SEP-2000) Genoscope - Centre National de Sequencage, 2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail : segrefgenoscope.cns.fr - Web : www.genoscope.cns.fr)

COMMENT

This GSS is part of a random genomic sequencing program of thirteen yeast species: *Saccharomyces bayanus* var. *uvatum*, *Saccharomyces exiguus*, *Saccharomyces servazzii*, *Zygosaccharomyces rouxii*, *Saccharomyces kluyveri*, *Kluyveromyces thermotolerans*, *Kluyveromyces lactis* var. *lactis*, *Kluyveromyces marxianus* var. *marxianus*, *Pichia angusta*, *Debaryomyces hansenii* var. *hansenii*, *Pichia sorbitophila*, *Candida tropicalis* and *Yarrowia lipolytica*. Genomic inserts of 3 to 5 kb were prepared and both extremities were sequenced. See keywords for description of this sequence and for the sequence of the other extremity of this insert.

FEATURES

Source

Location/Qualifiers
 1..985

Db

607 ACTGCAAGTTCTACGAANTGCTGCTGATCTGCTACATTTATGCACTTTTCAACAGTT 666

Oy

1014 cactccattggaagcgtgtaacttgaataaacttgaagttgagccgtagagttgg 1073

Db

667 CACGGCATTGAGGAGGCGGAACCTACAGAGCTGATCGAGGCTTGAGCTTACCGGCTGG 726

Oy

1074 gtccctctagataagatttgcattatgtagcgtcgaacaaagattctcattcaaacgc 1133

Db

727 CTTCACACTGCATCAAAATTTGCTGTGTGAGCGGCTGCACAAAGCTTATCCCATCAAGC 786


```

RESULT      4
AC2923982   328 bp   DNA      GSS      01-APR-2001
LOCUS       4906.1b97h11.s1 Saccharomyces mikatae IFO 1815 Saccharomyces
DEFINITION  mikatae genomic clone 4906.1b97h11.s1, DNA sequence.
ACCESSION   A2923982
VERSION     A2923982.1 GI:13494880
KEYWORDS
SOURCE      Saccharomyces mikatae.
ORGANISM    Saccharomyces mikatae.
REFERENCE   1 (bases 1 to 328)
AUTHORS     Cliften,P.F., Hillier,L.W., Fulton,L., Graves,T., Miner,T., Gish
            ,W.R., Waterston,R.H. and Johnston,M.
TITLE       Surveying Saccharomyces genomes to identify functional elements by
            comparative DNA sequence analysis
            Unpublished (2001)
JOURNAL
COMMENT      Department of Genetics
            Washington University Medical School
            Box 8232, 4566 Scott Ave., St. Louis, MO 63110, USA
            Tel: 314 362 2735
            Fax: 314 362 7855
            Email: mjgenetics.wustl.edu
            Class: random plasmid subclone.
FEATURES
    source          1..328
                    /organism="Saccharomyces mikatae"
                    /strain="IFO 1815"
                    /db_xref="taxon:114525"
                    /clone="4906.1b97h11.s1"
                    /clone_11b="Saccharomyces mikatae IFO 1815"
                    /note="Random genomic sequence"
BASE COUNT      93 a      66 c      54 g      115 t
ORIGIN
Query Match      14.2% Score 259.8; DB 13; Length 328;
Best Local Similarity 87.2%; Pred. No. 5.4e-34;
Matches 285; Conservative 0; Mismatches 42; Indels 0; Gaps 0;

OY 1102 gacgcctcaaaaagatcttcacatcaacacatattcacaggttcacaccc 1161
    |||||||
DB 2 GATGGCTCAAAAGATCATCTCATTCGAAAGCATATTTACAGGGTTACATTACACT 61

OY 1162 aagagaattgttctgcagacattagtaacagtaattctactgataaattacgct 1221
    |||||||
DB 62 AAGAGAAATTGTTTATTGATCTTGTAGTAACTAATCCACCGATGAATTTACAGCT 121

OY 1222 gtttggccatgaataatgcgtacgcgcaaaaacacatctgtaataatgctcctt 1281
    |||||||
DB 122 GTTTGGCTCATGAATGCGTACGCGCAAGAAAACACATTGTTAATGTGTCATCTTT 181

OY 1282 agtcaattgacaccttcacatttcctcccttccacagcattacagaataacata 1341
    |||||||
DB 182 AGTCAATTGACACACTTCTCTTATTTCTCTCTTTTACACAGATATCTACAGAACATCAGC 241

OY 1342 tttaacaacacctgcgcttctcttagagaagaagtcacactgcagtttgcgtccgt 1401
    |||||||
DB 242 TTCTCAACACCTTTGGCTTTTGTAGAGAGTCTAGACAGTGTGCTGATCGGTT 301

OY 1402 atcactaaggaattcccatatcat 1428
    |||||||
DB 302 ATCACTAAAGAAATCCCAATTATCAAT 328

```

```

ACCESSION   AL392570
VERSION     AL392570.1 GI:12142061
KEYWORDS    GSS.
SOURCE      Zygosaccharomyces rouxii.
ORGANISM    Zygosaccharomyces rouxii.
REFERENCE   1 (bases 1 to 893)
AUTHORS     de Montigny,J., Straub,M., Potier,S., Tekala,F., Dujon,B.,
            Wincker,P., Artiguenave,F. and Souclet,J.
TITLE       Genomic exploration of the hemiascomycetous yeasts : 8.
            Zygosaccharomyces rouxii
JOURNAL     FEBS Lett. 487 (1), 52-55 (2000)
MEDLINE
REFERENCE   20584718
AUTHORS     2 (bases 1 to 893)
            Souclet,J.L., Aigle,M., Artiguenave,F., Blandin,G.,
            Boloitin-Fukuhara,M., Bon,E., Brothier,P., Casaregola,S.,
            de-Montigny,J., Dujon,B., Durrens,P., Lepingle,A., Lorente,B.,
            Malpertuy,A., Neuvéglise,C., Ozier-Kalogeropoulos,O., Potier,S.,
            Saurin,W., Tekala,F., Toffano-Nicolas,C., Wesolowski-Louvel,M.,
            Wincker,P. and Weissenbach,J.
TITLE       Genomic exploration of the hemiascomycetous yeasts : 1. A set of
            yeast species for molecular evolution studies
JOURNAL     FEBS Lett. 487 (1), 3-12 (2000)
MEDLINE
REFERENCE   20584711
AUTHORS     3 (bases 1 to 893)
            Genoscope.
TITLE       Direct Submission
JOURNAL     Submitted (06-SEP-2000) Genoscope - Centre National de Séquençage,
            2 rue Gaston Crémieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :
            sequef@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT      This GSS is part of a random genomic sequencing program of thirteen
            yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces
            exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii,
            Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces
            lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia
            angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila,
            Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to
            5 kb were prepared and both extremities were sequenced. See
            keywords for description of this sequence and for the sequence of
            the other extremity of this insert.
FEATURES
    source          1..893
                    /organism="Zygosaccharomyces rouxii"
                    /strain="CBS 732"
                    /db_xref="taxon:4956"
                    /clone="AROAU05F04"
                    /clone_11b="AR0A"
                    /note="end : 77"
                    <2..>547
                    /note="similar to Saccharomyces cerevisiae ORF YJR117" [
                    SRE24 : zinc metallo-protease ]
                    /evidence=not_experimental
BASE COUNT      245 a      167 c      179 g      300 t      2 others
ORIGIN
Query Match      13.1% Score 238.8; DB 13; Length 893;
Best Local Similarity 65.5%; Pred. No. 1.3e-30;
Matches 368; Conservative 1; Mismatches 184; Indels 9; Gaps 1;

OY 1162 aagagaattgttctgcacattagtaacagtaattctactgataaattacgct 1221
    |||||||
DB 5 AAACGATTGTACTTATGATATCTTGGTCAATGAGAGTTCAGTGAAGAAATTTACGGCA 64

OY 1222 gtttggccatgaataatgcgtacgcgcaaaaacacatctgtaataatgctcctt 1281
    |||||||
DB 65 GTGCTGGCACATGAATATGTCATTCGCAAAAAGAACACATGTCATGCTGTATG 124

OY 1282 agtcaattgacaccttcacatttcctcccttccacagcattacagaataacata 1341
    |||||||
DB 125 AGTCAAGTTCAATCTTGTGATTAATTCATTCACGCGCAGTTTATAGAAATTTGCA 184

```


[illegible]

| RESULT | 6 |
|------------|---|
| A2930382 | |
| LOCUS | A2930382 517 bp DNA |
| DEFINITION | GSS 01-APR-2001 474.dhz53f03.sl <i>Saccharomyces unisporus</i> NRRL Y-1556 <i>Saccharomyces unisporus</i> genomic clone 474.dhz53f03.sl, DNA sequence. |

| | | |
|----------|-------------------------|-------------|
| VERSION | AZ930382.1 | GI:13501291 |
| KEYWORDS | GSS. | |
| SOURCE | Saccharomyces unisporus | |

ORGANISM Saccharomyces unisporus
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; Saccharomycetaceae; Saccharomyces.

| REFERENCE | AUTHORS |
|--------------------|---|
| 1 (bases 1 to 517) | Cliften, P.F., Hillier, L.W., Fulton, L., Graves, T., Miner, T., Gish |

TITLE ,W.R., Waterston,R.H. and Johnston,M.
Surveying Saccharomyces genomes to identify functional elements by
comparative DNA sequence analysis
JOURNAL Genetics 160 (2001)

JOURNAL
UNPUBLISHED (2001)
COMMENT
Contact: Johnston M

Department of Genetics
Washington University Medical School
Box 8232, 4566 Scott Ave., St. Louis, MO 63110, USA

Tel: 314 362 2735
Fax: 314 362 7855

Email: mj@genetics.wustl.edu.
Class: random plasmid subclone

| FEATURES | Location/Qualifiers |
|----------|---------------------|
| Courses | 1 517 |

```
1. .51/  
/organism="Saccharomyces unisporus"  
source
```

```

/strain="NRRL Y-1556 (CE
/db_xref="taxon:27294"

```

```
/clone="474.dhz53f03.s1"  
/clone_1fb="Saccharomyces uiiisnorus NBRI Y-1556"
```

```

PAGE COLUMN 153 73 50 313
/note="Random genomic sequen

```

| | | | | | | |
|------------|-----|---|----|---|-----|---|
| BASE COUNT | 132 | A | 19 | G | 213 | T |
| ORIGIN | . | | | | | |

| | | | | |
|-----------------------|--------|--------------------|--------|-------------|
| Query Match | 11.7%; | Score 213.6; | DB 13; | Length 517; |
| Best Local Similarity | 63.4%; | Pred. No. 2.4e-26; | | |

Matches 327; Conservative 0; Mismatches 189; Indels. ∴ 0; Gaps 0.

454 taccagaagctatctgaacaagttgccactgtgtctggaagacgaattgatgatgaa 513

| | | | |
|----|-----|--|-----|
| Db | 2 | TACAGGTTTATCCAGAAACATTACCTCCAGSTTCTTGAGGTGAATTTGACMAGMA | 61 |
| OY | 514 | actttcataaatcaaggaactactcccggyccaaagccaaagtlctccatttcoglygac | 573 |
| Db | 62 | ACTTTGAAAAAATCGAAGAATATTCCAAAGCTFAAGATTAAGTCTCATGCTCTGAT | 120 |
| OY | 574 | glctataactagcccaaaagctagttttcataaaatagaacctctccctaaatcttg | 633 |
| Db | 122 | ACCATCGGTTTACTCCAAAAACATTGCTTTATTAATAATGATATATCCAAAAATTATGG | 181 |
| OY | 634 | cacatggcgglttcttcatlgaatgcagtcctgcagtcagatctcatlgtctccac | 693 |
| Db | 182 | AATTTTGTCATATGTATCTGGAAGCTGTCTCCCTCTAAATATCAAGCTGTCTATT | 241 |
| OY | 694 | gtcgacagagtltaagctctcttggtgctctctatccagtlgtctacctgtytgaltg | 753 |
| Db | 242 | GTTGCTCAAAAGTTTGAGTGTTTTATTCATATGCTCAATTATATTCACGTGTTTAGCCTTA | 301 |
| OY | 754 | ccactctcttactatagccatltgtctcggaagaaaaatttgtttccaataatgacc | 813 |
| Db | 302 | CCAAATTCCTATTAATTTCTCATATTCGTCTTGAAGAAAAAATTTGGTTTAAATAGTAATCT | 361 |
| OY | 814 | gtccaactatgagatccaccgatatgatcaagaagtcgaacttggcgatatgtatgtygac | 873 |
| Db | 362 | GTTAAATTTATGAGATTATGACACATATTAAGGTAAATTTGTTAGGTTATGTGATTTGGTGT | 421 |
| OY | 874 | ccaactcttactcgttccctaaagacttggataaattccctactgaattctccttggtaac | 933 |
| Db | 422 | CCAATTCCTTATTAATTTCTTAAGAAATTTTGATTAATTTGAACATAATTTCTTATGCTAT | 481 |
| OY | 934 | attaagctctctcgttcgttcgttcgccaactcttgacc | 969 |
| Db | 482 | AATTTGTTAATTTTCTTTGTTGTCAGATGTTTACC | 517 |

RESULT 7
CNS06CRM/C

| LOCUS | 1116 bp | DNA |
|----------|---------|-----|
| CNS06CBM | | |

DEFINITION
1 / end of clone AR06AA003F08 of library
Zygosaccharomyces rouxii, genomic surv

ACCESSION AL392312 GI:12141541
VERSION AL392312.1

KEYWORDS GSS. 2v0205accharomvces rouyji SOURCE

ORGANISM
Zygosaccharomyces rouxii

Eukaryota; Fungi; Ascomycota; Sacccharomycetales; Sacccharomycetaceae;

| | | |
|-----------|--|---|
| REFERENCE | 1 (bases 1 to 1116) | : |
| AUTHORS | de Montigny, J., Straub, M., Potier, S., | |

Wincker, P., Artiguenave, F. and Souclet
Genomic exploration of the hemiascomycete

genomic exploitation of the nematocyst
zygocacharomyces rouxi
1970-1974 407-413 50-55 100000

JOURNAL
FEBES Lett. 48/ (1), 52-55 (2000)
MEDLINE
20584718

| | |
|-----------|---|
| REFERENCE | 2 (bases 1 to 1116) |
| AUTHORS | Souci ^{et} , J. L., Aigle, M., Artiquenave, F. |

Bolotin-Fukuhara, M., Bon, E., Brottier,
de-Montigny, J., Dutoit, B., Durand, P.

Malpertuy, A., Neuveglise, C., Ozier-Kal

Saurin, W., Tekala, F., Toffano-Nioche, C.
Wincker, P. and Weissenbach, J.

TITLE Genomic exploration of the hemiascomycetous yeast species for molecular evolution

JOURNAL
OF
FEBES Lett. 487 (1), 3-12 (2000)
MEDICINE
20584711

3 (bases 1 to 1116)

| AUTHORS | genoscope. |
|---------|-------------------|
| TITLE | Direct Submission |

JOURNAL
Submitted (06-SEP-2000) Genoscope - Ce
2 rue Gaston Cremieux, CP 5706, 91057

segreif@genoscope.cns.fr - Web : www.genoscope.cns.fr
 This CSS is part of a random genomic

Db 560 CTTTCGCCCTTACTTACTGCTTTGTTGTTCTCCACATGCTTATGTACGCCG 501
 QY 980 ttcaagctcatcatccatccatgttaataagttcaacccatcttgaggaggtgaacga 1039
 Db 500 TTCTACTCTGATCATGCGCCATGTTCAACAAGTTGAAGACTTGAGAGCGCGAGCTCA 441
 QY 1040 aaaaactatctgaagtttgccgatagagttgggtccctctagataagattttgtca 1039
 Db 440 AGAGCCGATCGGAGGCTCTTCCCAAGAAGCTGAGACTTCCCTTTCGACCTTTATGCA 381
 QY 1100 ttgacgctcaaaaagatcttctcatcaaacgatatctacaggttgccatccact 1159
 Db 380 TTGAGCGCTTACCCGATCCGCTCAGTCCATGCAATGCTTCTTCCGCTTGGGTTGGAAG- 320
 QY 1160 ccaagagaatgtttgttgacacatttgagaaacagaaatctacatgaataaggg 1219
 Db 321 -AAGCAGATTGCTCTGACGACACTGTGATGAGCAATGCTCCACAGCAGATTGAGG 264
 QY 1220 ctgtttggccatgaatcgctacgtgcaaaaaaacacatcgtaataatgtcatct 1279
 Db 263 CCATCTTGCGCCAGAGCTGCTGATCGAAGATGAAACATATCTCTGACAGCGCTGCTG 204
 QY 1280 ttatgaattgcaacacctctcatcttccctcttaccagacatctacagaatacat 1339
 Db 203 CCGGCAACGCAACATTTCTGACTGTGACTTGTGATTCCTGCGTTCTCCACAGCACT 144
 QY 1340 catttacaacaccttggtcttttc 1365
 Db 143 CTTTTCACCTTMCCTGCGCTTCTTC 118

RESULT 9

LOCUS B6176721 859 bp mRNA EST 06-FEB-2001
 DEFINITION 60231353F1 NIH_MGC_85 Homo sapiens cDNA clone IMAGE:4419344 5',
 mRNA sequence.
 ACCESSION B6176721
 VERSION B6176721.1 GI:12683424
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Cetartihii; Homnidae; Homo.

REFERENCE 1 (bases 1 to 859)
 AUTHORS NIH-MGC http://mgc.nci.nih.gov/
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgabs-remail.nih.gov
 Tissue Procurement: Louis Staudt, M.D., Ph.D.
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at:
 http://image.llnl.gov
 Plate: LLM10155 row: a column: 09
 High quality sequence stop: 492.
 Location/Qualifiers
 1..859

FEATURES

SOURCE

/organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_image="4419344"
 /clone_id="NIH_MGC_85"
 /tissue_type="lymphoma, cell line"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: lymph. Vector: PCMV-SPORT6, site:1; Not:
 site:2; Sali: cloned unidirectionally; oligo-df primed.
 Average insert size 1.867 kb. Library enriched for
 full-length clones and constructed by Life Technologies.
 Note: this is a NIH_MGC Library."
 BASE COUNT 214 a 209 c 171 g 265 t

ORIGIN

Query Match 5.9% Score 108.4; DB 11; Length 859;
 Best Local Similarity 54.6%; Pred. No. 7.3e-09;
 Matches 239; Conservative 0; Mismatches 196; Indels 3; Gaps 1;

QY 766 tatagccatttgcctcctgaagaanaatttggttccaataatgacccccaatgag 825
 Db 11 TATATACTTTTGTGATGAGAGAAAAACATGCTTCATCAACAGACTTTGGGTTCTTC 70
 QY 826 atccacgatgatgaagaagcttgacttggcgtatgtatgtgtggtgccaatcttac 885
 Db 71 ATGAAGATGATCAATCAAGAAATTTGTGTGACTCAGTGTATTTTGTCCGTGCTTCA 130
 QY 886 ctgttccttaagatcttggataaatccctactgattcccttggtaatatgtctc 945
 Db 131 CTTTACTTTACATTTATTAATAATGGGGGTACATTTTATTTATTTAGCTGCTGTTG 190
 QY 946 ttgttcgttgcacaatctttagccatgacatcatctcagttcatgcatgccaagtct 1005
 Db 191 ACATTAGTTGTGCTGCTGCTTCTTGTCAATCTATGCTGATTTATTTGCCCTTATTT 250
 QY 1006 aataagttcacctcattggaagcgttgaactgaaataatctatgaaagttggcgat 1065
 Db 251 GACAAATTCACACCTCTGCTGAGGAGAAAGCTTAAGAAGAAATTAAGTAATGCGAAG 310
 QY 1066 agagttgggtccctctagataagaatttggtaattgacgtcgtcaaaagatctctcat 1125
 Db 311 AGTATGACTTCTCTTACGACAGGCTGTATGCTGTGGAAGGATCTTAAACGCTCCAC 370
 QY 1126 tcaacgcatatttcacaggttggcattccctcaccacgaagaagatgttttgacact 1185
 Db 371 AGCAATGCTTATTTTATGAGCTT---CTTCAAGAACAAGCGACTATTTGCTTGACACT 427
 QY 1186 ttagtgaacagtaattct 1203
 Db 428 CTACTAGAAGAGTACTCT 445

RESULT 10

LOCUS AA965341 424 bp mRNA EST 31-JUN-1998
 DEFINITION e9b02a1.r1 Aspergillus nidulans 24hr asexual development and
 vegetative cDNA lambda zap library Emeritella nidulans cDNA clone
 e9b02a1 5', mRNA sequence.
 ACCESSION AA965341
 VERSION AA965341
 KEYWORDS EST.
 SOURCE Emeritella nidulans.

ORGANISM Emeritella nidulans.
 Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
 Eurotiiales; Trichocomaceae; Emeritella.
 REFERENCE 1 (bases 1 to 424)
 AUTHORS Kupfer, D., Gray, J., Hausner, J., Lai, H., Martin, W., Aramayo, R.,
 Prade, R. and Roe, B.
 TITLE An Aspergillus nidulans EST Database
 JOURNAL Unpublished (1998)
 COMMENT Contact: Bruce A. Roe, University of Oklahoma, broe@ou.edu
 Department of Chemistry and Biochemistry
 Advanced Center for Genome Technology, University of Oklahoma
 620 Parington Oval, Norman, OK 73019, USA
 Tel: 405 325 4912
 Fax: 405 325 7762
 Email: broe@ou.edu
 We anticipate the future release of the cDNA clones to the Fungal
 Genetics Stock Center
 Seq primer: T3
 High quality sequence stop: 351.
 Location/Qualifiers
 1..424

FEATURES

SOURCE

/organism="Emeritella nidulans"
 /strain="FGSC A26"

/db_xref="taxon:162425"
 /clone="9b02a1"
 /clone_lib="Aspergillus nidulans 24hr asexual
 developmental and vegetative cDNA lambda zap library"
 /issue_type="vegetative mycelia, asexual structures"
 /note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
 XhoI; 5' end of cDNA cloned into EcoRI site of pBluescript
 3' end of cDNA cloned into XhoI site of pBluescript"
 BASE COUNT 96 a 99 c 100 g 129 t
 ORIGIN

Query Match 5.9%; Score 106.8; DB 10; Length 424;
 Best Local Similarity 53.3%; Pred. No. 1.7e-08;
 Matches 225; Conservative 0; Mismatches 197; Indels 0; Gaps 0;

Db 1 TTGAGGAGAAATTTGGCTTTAACAGACGAGGTTAACTGATGATCCGACATGCTTA 60
 Oy 782 ttgagaagaaatttggttcaataatgaccgtccaaactatgatacagatatgaca 841
 Db 1 TTGAGGAGAAATTTGGCTTTAACAGACGAGGTTAACTGATGATCCGACATGCTTA 60
 Oy 842 agagctgacttggcgatgctatgtgtgcccacacttacttcttcttaagatct 901
 Db 61 AGGCCAGATTTGGGATTTGTTCTGGCGGCCCATATTAGCGCCGTTTGAAGATTG 120
 Oy 902 ttgataaattccctactgatttcccttggtaacataatgctcttcttgcgttgcaca 961
 Db 121 TCCAGAAAGATGGGACCTTCATGCTTTTACTGCTTTGGCTTTTGGTCTTTGTTGACG 180
 Oy 962 tcttaagcaatgaacatcttccagcttcaatcattgcccacatgtttaataagttacatc 1021
 Db 181 TTTTCGCCATCACCATTACCTTACCGCATTTTGCCTTTTAAACAAAGCTTTCCTCC 240
 Oy 1022 tggagagacgtgaaactgaaaactatgaaagtttggcagatagagttgagttccctc 1081
 Db 241 TTGAACCTGGTGGCATCAAGACTGGCGTTGAGATCTTGCCAAAGACTCAACTTCCCGC 300
 Oy 1082 tagataagattttgtcatgacggtccaaaagaatcttccatccaacagcatattca 1141
 Db 301 TCCAGAACTCCATGATGATGATGAGCAGCGCAGCAGCAACAACTTATTTT 360
 Oy 1142 caggttgccttcaaccctcaagaagattgttgcgtcgaactttagtgaacgtaatt 1201
 Db 361 AGGCTTCCTCTTGAAGAAGACATTTGCTATTGATACCTTATGAGAAGAGTGAGC 420
 Oy 1202 ct 1203
 Db 421 CT 422

RESULT 11
 BF051006 543 bp mRNA EST 18-MAY-2001
 LOCUS BF051006
 DEFINITION esculentum cDNA clone cLEM21E12 5' sequence, mRNA sequence.
 ACCESSION BF051006
 VERSION BF051006.1 GI:10804902
 KEYWORDS EST.
 SOURCE tomato.
 ORGANISM Lycopersicon esculentum
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Asteridae; euasterids I; Solanales; Solanaceae; Solanum;
 Lycopersicon.
 1 (bases 1 to 543)
 Alcala,J., Vredolov,J., White,R., van der Hoeven,R.S., Holt,I.E.,
 Liang,F., Hansen,T.S., Craven,M.B., Bowman,C.L., Romling,C.M.,
 Nierman,W., Fraser,C.M., Martin,G.B., Giovannoni,J.D. and Tanksley
 S.D.
 TITLE Generation of ESTs from tomato fruit tissue, immature green
 JOURNAL Unpublished (2000)
 COMMENT Contact: CUGI
 Clemson University Genomics Institute
 Clemson University

100 Jordan Hall, Clemson, SC 29634, USA
 Email: <http://www.genome.clemson.edu/orders/index.html>.
 Location/Qualifiers
 1. 543
 /organism="Lycopersicon esculentum"
 /cultivar="T496"
 /db_xref="taxon:4081"
 /clone="cLEM21E12"
 /clone_lib="tomato developing/immature green fruit"
 /issue_type="fruit"
 /dev_stage="immature green (5-35 days post-anthesis)"
 /lab_host="SOLR"
 /note="Vector: pBluescriptSKmCunadap; Site_1: EcoRI;
 Site_2: XhoI; Fruit were tagged at 5 dpa (0.5 cm) and
 harvested at 7 day intervals through 35 dpa. Equal masses
 of tissue from each stage were combined (including seeds
 and locules) prior to mRNA isolation."

BASE COUNT 143 a 113 c 116 g 171 t
 ORIGIN

Query Match 5.8%; Score 105.6; DB 11; Length 543;
 Best Local Similarity 50.0%; Pred. No. 2.5e-08;
 Matches 264; Conservative 0; Mismatches 264; Indels 0; Gaps 0;

Oy 732 ttgtctcaacttggtagtggccactcttactatgaccatttgcctggagaagaaa 791
 Db 16 TTGGTCACAGATCACCGATCTGCAATTTCTTGATTAACAACCTTTGATTCAGGCGCG 75
 Oy 792 attggttcaataatgacgcctcaacttggatcagcagatagatcaagaatctgac 851
 Db 76 CCATGGGTTTCAACAAGCAAGAGTGGTGTACTTTCAGGATATGATMAAAGAAATTC 135
 Oy 852 ttggcgtatgtatttgggtggccaatcttactcgttcttcaagatcttgaataatt 911
 Db 136 TTTATCTATTGTGATGTGCTCCCATTTGTGCTGCCATTTATGTATAGTCAGAAAGC 195
 Oy 912 cccctactgattccttctgtatcatatgcttcttcttgcgttgcacaacttagcat 971
 Db 196 AGGTCTTACTTGGCAATTTACTTATGAGGATTTATGCTGTGTTGGCTCTTGTATGAT 255
 Oy 972 gacaatcattccagctctcattcattccatgltttaataagttcaactccattgagagcg 1031
 Db 256 GACTATCTACCCAGACTAATTTGCTCCACTTTTCAACAAGATTCACCTCCGTTCCACAGG 315
 Oy 1032 tgaactgaaaaactatgaaagtltggccgataagatgttggttccctcagaataat 1091
 Db 316 AGACCTTAGATTGAAGATCGAAGATCTTGCAATCTCACTCAAAATTTCCCTTAAAGAAAT 375
 Oy 1092 ttgtcatgagcgctcaaaaagatcttccatccaagcagatatccagggttggc 1151
 Db 376 GTTTGTTGTTGATGAGCTCCACAGGTCACAGCATATGATCATATGATGGAATTC 435
 Oy 1152 attacctccaagaagtgttgcgtgacaactttagtgaacgtaattctactatgta 1211
 Db 436 CAAGAACAGCGCATTTGCTCTATGATATACATCATCAGCAGAGCAAGATGATGAGGA 495
 Oy 1212 aattaagcgtgttggcccatgaaatcgttcaactgtgcacaaataaaca 1259
 Db 496 AATGCTTGCTGTTATTTGACACAGAGCTTGCTGCTGCAAACTTATATCA 543

RESULT 12
 AM174887 708 bp mRNA EST 07-JUN-2001
 LOCUS AM174887
 DEFINITION f606101.y1 zebrafish wasnu MPIMG EST Danio rerio cDNA clone
 IMAGE:3738097 5' similar to TR:075844 075844 HS STE24P.;, mRNA
 sequence.
 ACCESSION AM174887
 VERSION AM174887.1 GI:6440886
 KEYWORDS EST.
 SOURCE zebrafish.
 ORGANISM Danio rerio

Tue Feb 12 16:22:23 2002

us-09-165-460a-1.rst

Page 11

| Accession | Sequence | Length |
|-----------|---|--------|
| Db | 577 ACGGAGAGCTAAAAAAGACATTCACAGTCTGGCAGACAGTGTGAATTTCTTTGGGCC | 518 |
| Oy | 1088 agatttttgcatgacggtcgaagaagatctctcatcacaagatattcacaagt | 1147 |
| Db | 517 AACTTTGTGCTGCAAGGTTCAAGCGTTCTCGCATAGTAAAGGTTACTTTACCGAT | 458 |
| Oy | 1148 tg 1149 | |
| Db | 457 TG 456 | |

Search completed: February 12, 2002, 12:01:41
Job time: 6514 sec

THIS PAGE BLANK (25PT0)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 12, 2002, 11:09:13 : Search time 91.73 Seconds

(without alignments)
4505.858 Million cell updates/sec

Title: US-09-165-460a-1

Perfect score: 1825

Sequence: 1 accatactttttctatct.....aaataaacgattaacatt 1825

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 351203 seqs, 113238999 residues

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents_NA:*

1: /cgn2_6/ptodata/2/ina/5A.COMB.seq:*

2: /cgn2_6/ptodata/2/ina/5B.COMB.seq:*

3: /cgn2_6/ptodata/2/ina/6A.COMB.seq:*

4: /cgn2_6/ptodata/2/ina/6B.COMB.seq:*

5: /cgn2_6/ptodata/2/ina/PCBUS.COMB.seq:*

6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|-------------------|
| 1 | 136 | 7.5 | 686 | 4 | US-08-998-416-929 |
| 2 | 117.4 | 6.4 | 2968 | 3 | US-09-022-699-1 |
| 3 | 54.6 | 3.0 | 7218 | 1 | US-08-232-463-14 |
| 4 | 51.6 | 2.8 | 1641 | 1 | US-08-300-903A-8 |
| 5 | 49 | 2.7 | 1368 | 3 | US-08-874-563-5 |
| 6 | 49 | 2.7 | 1368 | 3 | US-08-577-483-14 |
| 7 | 48 | 2.6 | 8220 | 2 | US-08-568-459A-11 |
| 8 | 48 | 2.6 | 8220 | 2 | US-08-487-826B-11 |
| 9 | 48 | 2.6 | 19124 | 2 | US-08-487-826B-13 |
| 10 | 47.8 | 2.6 | 2035 | 3 | US-09-022-699-3 |
| 11 | 47.4 | 2.6 | 1016 | 1 | US-08-109-391A-3 |
| 12 | 47.4 | 2.6 | 1016 | 1 | US-08-459-019A-3 |
| 13 | 47.4 | 2.6 | 1016 | 1 | US-08-460-428A-3 |
| 14 | 47.4 | 2.6 | 1016 | 3 | US-08-458-860A-3 |
| 15 | 47.4 | 2.6 | 4253 | 3 | US-08-577-483-7 |
| 16 | 47 | 2.6 | 870 | 2 | US-08-420-629-9 |
| 17 | 46.8 | 2.6 | 4254 | 4 | US-08-443-639-7 |
| 18 | 46.8 | 2.6 | 80246 | 4 | US-09-078-294-4 |
| 19 | 46.2 | 2.5 | 1733 | 3 | US-09-073-569-1 |
| 20 | 46.2 | 2.5 | 6727 | 3 | US-08-629-643A-5 |
| 21 | 46.2 | 2.5 | 6727 | 4 | US-09-280-799-1 |
| 22 | 46.2 | 2.5 | 6727 | 4 | US-09-155-884-5 |
| 23 | 46.2 | 2.5 | 7218 | 1 | US-08-232-463-14 |
| 24 | 46 | 2.5 | 2082 | 2 | US-08-785-310A-2 |
| 25 | 46 | 2.5 | 2447 | 2 | US-09-014-969-14 |
| 26 | 45.8 | 2.5 | 144 | 1 | US-08-702-344-26 |
| 27 | 45.8 | 2.5 | 240 | 1 | US-08-628-417-6 |

| | | | | | | |
|----|------|-----|-------|---|-------------------|--------------------|
| 28 | 45.6 | 2.5 | 1051 | 4 | US-09-245-041-10 | Sequence 10, Appl |
| 29 | 45 | 2.5 | 1890 | 6 | 5312912-3 | Patent No. 5312912 |
| 30 | 45 | 2.5 | 3072 | 4 | US-09-522-217-55 | Sequence 55, Appl |
| 31 | 44 | 2.4 | 1117 | 4 | US-08-247-373B-33 | Sequence 33, Appl |
| 32 | 44 | 2.4 | 8920 | 2 | US-08-446-855A-1 | Sequence 1, Appl |
| 33 | 44 | 2.4 | 8920 | 4 | US-09-150-741-1 | Sequence 1, Appl |
| 34 | 43.8 | 2.4 | 17949 | 4 | US-09-087-465-3 | Sequence 3, Appl |
| 35 | 43.4 | 2.4 | 2551 | 1 | US-08-486-013-70 | Sequence 70, Appl |
| 36 | 43.4 | 2.4 | 2551 | 2 | US-08-482-279-70 | Sequence 70, Appl |
| 37 | 43.4 | 2.4 | 2551 | 2 | US-08-342-268-70 | Sequence 70, Appl |
| 38 | 43.4 | 2.4 | 2551 | 2 | US-09-015-868-70 | Sequence 70, Appl |
| 39 | 43.4 | 2.4 | 2551 | 4 | US-09-397-386-70 | Sequence 70, Appl |
| 40 | 43.4 | 2.4 | 4718 | 3 | US-08-936-135-9 | Sequence 9, Appl |
| 41 | 43.4 | 2.4 | 4733 | 3 | US-08-936-135-11 | Sequence 11, Appl |
| 42 | 43.4 | 2.4 | 4769 | 3 | US-08-936-135-13 | Sequence 13, Appl |
| 43 | 43.4 | 2.4 | 4784 | 3 | US-08-936-135-15 | Sequence 15, Appl |
| 44 | 43.2 | 2.4 | 300 | 1 | US-08-222-177A-45 | Sequence 45, Appl |
| 45 | 43.2 | 2.4 | 3481 | 4 | US-08-965-729A-1 | Sequence 1, Appl |

ALIGNMENTS

RESULT 1
US-08-998-416-929/c
Sequence 929, Application US/08998416
Patent No. 6239264
GENERAL INFORMATION:
APPLICANT: Philippesen, Peter
APPLICANT: Pohmann, Rainer
APPLICANT: Steiner, Sabine
APPLICANT: Mohr, Christine
APPLICANT: Wendland, Jurgen
APPLICANT: Knechtel, Philipp
APPLICANT: Reibschung, Corinne
TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSSPYII
TITLE OF INVENTION: AND USES THEREOF
NUMBER OF SEQUENCES: 1152
CORRESPONDENCE ADDRESSES:
ADDRESSEE: No. 6239264artis Corporation
STREET: 3054 Cornwallis Road
CITY: Research Triangle Park
STATE: No. 6239264th Carolina
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/998,416
FILING DATE: 24-DEC-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: CH 0016/97
FILING DATE: 31-DEC-1996
ATTORNEY/AGENT INFORMATION:
NAME: Meigs, J. Timothy
REGISTRATION NUMBER: 38,241
REFERENCE/DOCKET NUMBER: PF/5-30306/A/CGC1976
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8587
TELEFAX: 919-541-8689
INFORMATION FOR SEQ ID NO: 929:
SEQUENCE CHARACTERISTICS:
LENGTH: 686 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: PAG1577RP


```

1      TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
2      TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
3      NUMBER OF SEQUENCES: 45
4      CORRESPONDENCE ADDRESS:
5      ADDRESSEE: Knobbe Martens Olson & Bear
6      STREET: 620 Newport Center Drive 16th Floor
7      CITY: Newport Beach
8      STATE: California
9      COUNTRY: US
10     ZIP: 92660
11     COMPUTER READABLE FORM:
12     MEDIUM TYPE: Floppy disk
13     COMPUTER: IBM PC compatible
14     OPERATING SYSTEM: PC-DOS/MS-DOS
15     SOFTWARE: PatentIn Release #1.0, Version #1.25
16     CURRENT APPLICATION DATA:
17     APPLICATION NUMBER: US/08/487,826B
18     FILING DATE: 10-SEP-1993
19     CLASSIFICATION: 435
20     ATTORNEY/AGENT INFORMATION:
21     NAME: Israelson, Ned
22     REGISTRATION NUMBER: 29,655
23     REFERENCE/DOCKET NUMBER: NIH121.001CP1
24     TELECOMMUNICATION INFORMATION:
25     TELEPHONE: (619) 235-8550
26     TELEFAX: (619) 235-0176
27     INFORMATION FOR SEQ ID NO: 11:
28     SEQUENCE CHARACTERISTICS:
29     LENGTH: 8220 base pairs
30     TYPE: nucleic acid
31     STRANDEDNESS: single
32     TOPOLOGY: linear
33     MOLECULE TYPE: DNA (genomic)
34     HYPOTHEITICAL: NO
35     ORIGINAL SOURCE:
36     ORGANISM: Plasmodium falciparum
37     US-08-487-826B-11
38
39     Query Match 2.6%; Score 48; DB 2; Length 8220;
40     Best Local Similarity 59.6%; Pred. No. 0.027;
41     Matches 81; Conservative 0; Mismatches 55; Indels 0; Gaps 0;
42
43     QY 20 ttcaacaagaaagccttacacacacacacacatcatctacatcatcacacaata 79
44         || || || || || || || || || || || || || || || || || || ||
45     Db 8019 TACATATACAGAAATCATATACATACACATATATATATATCCAAACACATACATATA 7960
46
47     QY 80 tacataatgtaaactgtatattcattccctataaccacaaagaggaacttaacttt 139
48         || || || || || || || || || || || || || || || || || || ||
49     Db 7959 CATATATATATATATATATATATATCCAAACATATACCAAAATACATATACATATACAT 7900
50
51     QY 140 cccctctttctcagct 155
52         || || || || || || || || || || || || || || || || || || ||
53     Db 7899 CCACATACATCCACAT 7884
54
55     RESULT 9
56     US-08-487-826B-13/C
57     Sequence 13; Application US/08487826B
58     Patent No. 5993827
59     GENERAL INFORMATION:
60     APPLICANT: Sim, Kim L.
61     APPLICANT: Chitluis, Chetan
62     APPLICANT: Miller, Louis H.
63     APPLICANT: Peterson, David S.
64     APPLICANT: Su, Xin-zhaun
65     APPLICANT: Wellens, Thomas E.
66     TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
67     TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
68     NUMBER OF SEQUENCES: 45
69     CORRESPONDENCE ADDRESS:
70     ADDRESSEE: Knobbe Martens Olson & Bear
71     STREET: 620 Newport Center Drive 16th Floor

```

RESULT 11
US-08-109-391A-3/C
Sequence 3, Application US/08109391A
Patent No. 5639876
GENERAL INFORMATION:
APPLICANT: tripp, Cynthia A.
APPLICANT: Frank, Glenn R.
APPLICANT: Grileve, Robert B.
TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING NOVEL
TITLE OF INVENTION: PARASTITIC HELMINTH PROTEINS
NUMBER OF SEQUENCES: 17
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sheridan Ross & McIntosh
STREET: 1700 Lincoln St., Suite 3500
CITY: Denver
STATE: CO
COUNTRY: U.S.A.
ZIP: 80203
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/109,391A
FILING DATE: 19-AUG-1993
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Connell, Gary J.
REGISTRATION NUMBER: 32,020
REFERENCE/DOCKET NUMBER: 2618-13
TELECOMMUNICATION INFORMATION:
TELEPHONE: 303/863-9700

Patent No. 5912337
GENERAL INFORMATION:
APPLICANT: Tripp, Cynthia A.
APPLICANT: Frank, Glenn A.
APPLICANT: Grieve, Robert B.
TITLE OF INVENTION: NOVEL PARASITIC HELMINTH
TITLE OF INVENTION: P220 PROTEINS
NUMBER OF SEQUENCES: 17
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sheridan Ross P.C.
STREET: 1700 Lincoln St., Suite 3500
CITY: Denver
STATE: CO
COUNTRY: U.S.A.
ZIP: 80203
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/460,428A
FILING DATE: 02-JUN-1995
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Connell, Gary J.
REGISTRATION NUMBER: 32,020
REFERENCE/DOCKET NUMBER: 2618-13-3
TELECOMMUNICATION INFORMATION:
TELEPHONE: 303/863-9700
TELEFAX: 303/863-0223
INFORMATION FROM SEQ. ID NO.: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 1016 base pairs
type: nucleic acid

```

? STRANDNESS: single
? TOPOLOGY: linear
? MOLECULE TYPE: cDNA
? FEATURE:
? NAME/KEY: CDS
? LOCATION: 3..626
? FEATURE:
? NAME/KEY: 3'UTR
? LOCATION: 627..1016
?
US-08-460-428A-3

```

| | | | | | | | |
|-----------------------|-------|--------------|-------|------------|----|--------|------|
| Query Match | 2.6% | Score | 47.4 | DB | 2 | Length | 1016 |
| Best Local Similarity | 57.9% | Pred. No. | 0.022 | | | | |
| Matches | 84 | Conservative | 0 | Mismatches | 61 | Indels | 0 |
| | | | | | | Gaps | 0 |

[illegible]

RESULT 14
US-08-458-860A-3/c

; Sequence 3, Appl. No. 61003900;
; Patent No. 61003900

GENERAL INFORMATION: Frank, Glenn R.
APPLICANT: Trilp, Cynthia A.
APPLICANT: Grieve, Robert B.
TITLE OF INVENTION: NOVEL PARASITIC HELMINTH
TITLE OF INVENTION: P220 NUCLEIC ACID MOLECULES
NUMBER OF SEQUENCES: 17

CORRESPONDENCE ADDRESS: ;
ADDRESSEE: Sheridan Ross P.C. ;
STREET: 1700 Lincoln St., Suite 3500 ;
CITY: Denver ;

CITY: DENVER
STATE: CO
COUNTRY: U.S.A.
ZIP: 80203

```

215- 00205
?
? COMPUTER READABLE FORM:
?
? MEDIUM TYPE: Floppy disk
?
? COMPUTER: IBM PC compatible
?
? OPERATING SYSTEM: PC-DOS/MS-DOS
?
? SOFTWARE: PatentIn Release #1.0, Version #1.30

```

```

:
: CURRENT APPLICATION DATA:
:
: APPLICATION NUMBER: US/08/458,860A
:
: FILING DATE: 02-JUN-1995
:
:

```

; CLASSIFICATION: 536
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Connell, Gary J.
 ;

REGISTRATION NUMBER: 32,020
REFERENCE/DOCKET NUMBER: 2618-13-2
TELECOMMUNICATION INFORMATION:
TELEPHONE: 303/453-3700

TELEPHONE: 303/863-9700
TELEFAX: 303/863-0223
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS

```

; SEQUENCE CHARACTERISTICS:
; LENGTH: 1016 base pairs
; TYPE: nucleic acid
; CDNA: CDNA:1016

```

```

;          STRANDEDNESS: single
;          TOPOLOGY: linear
;          MOLECULE TYPE: cDNA
FEATURES:

```

```

;          FEATURE:
;          NAME/KEY:      CDS
;          LOCATION:      3..626

```

```

;         FEATURE:
;         NAME/KEY:      3'UTR
;         LOCATION:      627..1016
US-08-458-860A-3

```

| | | | | |
|--------------------------|-------|-----------------|----------|-------------|
| Query Match | 2.6% | Score 47.4 | DB 3 | Length 1016 |
| Best Local Similarity | 57.9% | Pred. No. 0.022 | | |
| Matches 84; Conservative | 0 | Mismatches 61 | Indels 0 | Gaps 0 |

| | | | |
|----|-----|--|------|
| OY | 1 | acactacccttttctcattcttcgaacaaagcgcccttagaacgcaacacctaactc | 60 |
| | | | |
| Db | 934 | ACATACATACATACATATTTACACATACCAACAACACACATACACACACACACAC | 87/5 |
| OY | 61 | acatacatatacaatatatacatatatgtfanaactigtatattcatlccatttaacaaa | 120 |
| | | | |
| Db | 874 | ACATACATACATACATATATATATATACAGAAATACTTATTGTTGAAGTATCATTTTAA | 81/5 |
| OY | 121 | aagaggcaattaactttccctcct | 145 |
| | | | |
| Db | 814 | AAAAAGAAAACCCTGTTTTTCATTAT | 790 |

RESULT 15
US-08-577-483-7/c

; Sequence 7, Application US/0857748;
; Patent No. 6100451
; SEQUENTIAL INFORMATION

GENERAL INFORMATION:
 APPLICANT: Chappell, Joseph
 APPLICANT: Yin, Shaohui
 APPLICANT: Cornett, Catherine A.G.
 TITLE OF INVENTION: Transcriptional Control Sequences and
 TITLE OF INVENTION: Methods
 NUMBER OF SEQUENCES: 14

;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Fish & Richardson, P.C.
;; STREET: 225 Franklin Street

```

; CITY: Boston
; STATE: MA
; COUNTRY: US

```

```

?      ZIP: 02110-2804
?      COMPUTER READABLE FORM:
?      MEDIUM TYPE: Floppy disk
?      COMPUTER: IBM PC compatible
?      OPERATING SYSTEM: PC-DOS/MS-DOS
?      SOFTWARE: PatentIn Release #1.0, Version #1.3

```

```

; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/577,48
; FILING DATE: 22-DEC-1995
; CLASSIFICATION: 000

```

CLASSIFICATION: 800
ATTORNEY/AGENT INFORMATION:
NAME: Paul T. Clark
DECISION NUMBER: 33 1

REGISTRATION NUMBER: 32,104
REFERENCE/DOCKET NUMBER: 07678/003000
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 542-5070

TELEPHONE: (617) 542-3
TELEFAX: (617) 542-8906
TELEX: 200154
INFORMATION FOR SEC ID NO:

SEQUENCE CHARACTERISTICS:
LENGTH: 4253 base pairs
TYPE: nucleic acid

STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA

ORIGINAL SOURCE.

ORGANISM: Nicotiana tabacum

```

; NAME/KEY: CDS
; LOCATION: 1216..132
; FEATURE:

```


THIS PAGE BLANK (23PT0)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 12, 2002, 09:25:34 ; Search time 40.57 Seconds
(without alignments)
827.094 Million cell updates/sec

Title: US-09-165-460A-2
Perfect score: 2342
Sequence: 1 MFDLKTIIDHPNIPMKLTIS.....HPTLAERSTALDYVSEKKKN 453

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues
Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
A_Geneseq_1101.*
1: /SID88/gcgdata/geneseq/geneseq/AA1980.DAT.*
2: /SID88/gcgdata/geneseq/geneseq/AA1981.DAT.*
3: /SID88/gcgdata/geneseq/geneseq/AA1982.DAT.*
4: /SID88/gcgdata/geneseq/geneseq/AA1983.DAT.*
5: /SID88/gcgdata/geneseq/geneseq/AA1984.DAT.*
6: /SID88/gcgdata/geneseq/geneseq/AA1985.DAT.*
7: /SID88/gcgdata/geneseq/geneseq/AA1986.DAT.*
8: /SID88/gcgdata/geneseq/geneseq/AA1987.DAT.*
9: /SID88/gcgdata/geneseq/geneseq/AA1988.DAT.*
10: /SID88/gcgdata/geneseq/geneseq/AA1989.DAT.*
11: /SID88/gcgdata/geneseq/geneseq/AA1990.DAT.*
12: /SID88/gcgdata/geneseq/geneseq/AA1991.DAT.*
13: /SID88/gcgdata/geneseq/geneseq/AA1992.DAT.*
14: /SID88/gcgdata/geneseq/geneseq/AA1993.DAT.*
15: /SID88/gcgdata/geneseq/geneseq/AA1994.DAT.*
16: /SID88/gcgdata/geneseq/geneseq/AA1995.DAT.*
17: /SID88/gcgdata/geneseq/geneseq/AA1996.DAT.*
18: /SID88/gcgdata/geneseq/geneseq/AA1997.DAT.*
19: /SID88/gcgdata/geneseq/geneseq/AA1998.DAT.*
20: /SID88/gcgdata/geneseq/geneseq/AA1999.DAT.*
21: /SID88/gcgdata/geneseq/geneseq/AA2000.DAT.*
22: /SID88/gcgdata/geneseq/geneseq/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match Length | ID | Description |
|------------|-------|--------------------|-----------|--------------------|
| 1 | 2317 | 98.9 | AAW48301 | Yeast CAAX process |
| 2 | 836 | 36.5 | AAAG15512 | Arabidopsis thalia |
| 3 | 855 | 36.5 | AAAG15513 | Arabidopsis thalia |
| 4 | 848 | 36.2 | AAAG15514 | Arabidopsis thalia |
| 5 | 698.5 | 29.8 | AAAY26896 | Human farnesylated |
| 6 | 653 | 27.9 | AAW67587 | Human metallopept |
| 7 | 571.5 | 24.4 | AAW58297 | Lung cancer associ |
| 8 | 439 | 19.6 | AAW20614 | H. pylori cell env |
| 9 | 454 | 19.4 | AAW46348 | H. pylori HPS187 p |
| 10 | 402.5 | 17.2 | AAW5578 | Human protein sequ |
| 11 | 402.5 | 17.2 | AAW86011 | Human CAAX process |

| | | | | | | |
|----|-------|------|------|----|----------|---------------------|
| 12 | 389 | 16.6 | 357 | 22 | AAW46414 | H. pylori HPC187 p |
| 13 | 328 | 14.0 | 426 | 20 | AAW05501 | Bacillus subtilis |
| 14 | 274.5 | 11.7 | 192 | 18 | AAW20512 | H. pylori cell env |
| 15 | 225.5 | 9.6 | 211 | 20 | AAW05502 | Bacillus subtilis |
| 16 | 150.5 | 6.6 | 116 | 20 | AAW05503 | Bacillus subtilis |
| 17 | 150.5 | 6.4 | 347 | 18 | AAW20747 | H. pylori inner me |
| 18 | 148.5 | 6.3 | 226 | 21 | AAW75220 | Neisseria gonorrhe |
| 19 | 148.5 | 6.3 | 226 | 21 | AAW75221 | Neisseria meningit |
| 20 | 148.5 | 6.3 | 226 | 21 | AAW75222 | Neisseria meningit |
| 21 | 143 | 6.1 | 192 | 18 | AAW20133 | H. pylori cell env |
| 22 | 140 | 6.0 | 155 | 20 | AAW36280 | Human secreted pro |
| 23 | 132 | 5.6 | 225 | 18 | AAW20328 | H. pylori inner me |
| 24 | 123.5 | 5.3 | 302 | 19 | AAW62687 | Streptococcus pneu |
| 25 | 109 | 4.7 | 841 | 21 | AAW11039 | S. xyloos mpef pr |
| 26 | 106 | 4.5 | 1226 | 18 | AAW13825 | Yeast transcrip |
| 27 | 104 | 4.4 | 333 | 22 | AAW87782 | Rat T2R02 amino ac |
| 28 | 103.5 | 4.4 | 185 | 22 | AAW73074 | Olfactory receptor |
| 29 | 103 | 4.4 | 471 | 20 | AAW20061 | B. burgdorferi ant |
| 30 | 103 | 4.4 | 490 | 20 | AAW20060 | B. burgdorferi ant |
| 31 | 101 | 4.3 | 378 | 21 | AAW10552 | Murine aspartate p |
| 32 | 100 | 4.3 | 542 | 21 | AAW44905 | Human potassium ch |
| 33 | 100 | 4.3 | 1102 | 21 | AAW33028 | Rat E1K1 potassium |
| 34 | 100 | 4.3 | 1107 | 21 | AAW44907 | Human potassium ch |
| 35 | 100 | 4.3 | 1228 | 22 | AAW01118 | Human prostate-spe |
| 36 | 100 | 4.3 | 1228 | 22 | AAW01118 | Human prostate-spe |
| 37 | 100 | 4.3 | 1261 | 22 | AAW01179 | Human prostate-spe |
| 38 | 100 | 4.3 | 1261 | 22 | AAW01179 | Human prostate-spe |
| 39 | 100 | 4.3 | 1325 | 20 | AAW33541 | Human MPR-relate |
| 40 | 99 | 4.2 | 314 | 13 | AAW27874 | Odorant receptor c |
| 41 | 98 | 4.2 | 314 | 13 | AAW27874 | Human olfactory re |
| 42 | 98 | 4.2 | 314 | 20 | AAW73010 | Olfactory receptor |
| 43 | 98 | 4.2 | 400 | 22 | AAW19850 | B. burgdorferi ant |
| 44 | 98 | 4.2 | 552 | 22 | AAW79559 | Cornebacterium gl |
| 45 | 98 | 4.2 | 581 | 22 | AAW92515 | C glutamincum prote |

ALIGNMENTS

| | | |
|----------|---|----------------------------|
| RESULT 1 | AAW48301 | standard; Protein: 453 AA. |
| ID | AAW48301 | |
| XX | AAW48301 | |
| AC | 20-JUL-1998 | (first entry) |
| XX | | |
| DT | | |
| XX | | |
| DE | | |
| XX | | |
| KW | Afc1p, Afc1 gene, a-factor convertase; CAAX protease; | |
| KW | zinc metalloprotease; yeast; prenylation; Ras; inhibitor; cancer; | |
| KW | colorectal carcinoma; pancreas carcinoma; leukaemia; therapy. | |
| XX | | |
| OS | Saccharomyces cerevisiae. | |
| XX | | |
| EH | Key | Location/Qualifiers |
| FT | Misc-difference 13 | /note= "encoded by ATC" |
| FT | Misc-difference 121 | /note= "encoded by AGT" |
| FT | Misc-difference 305 | /note= "encoded by AAC" |
| FT | Misc-difference 442 | /note= "encoded by TCG" |
| FT | | |
| XX | W09805786-A2. | |
| XX | | |
| PN | 12-FEB-1998. | |
| XX | | |
| PD | | |
| XX | | |
| PF | 06-AUG-1997; | 97WO-US14777. |
| XX | | |
| PR | 30-JUL-1997; | 97US-0902774. |
| PR | 07-AUG-1996; | 96US-0023491. |

PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139839.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145214.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.

PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151066.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161922.
PR 28-OCT-1999; 99US-0161992.
PR 29-OCT-1999; 99US-0162142.

Query Match 36.5%; Score 856; DB 21; Length 458;
Best Local Similarity 41.2%; Pred. No. 1,7e-82;

Matches 185; Conservative 73; Mismatches 167; Indels 24; Gaps 6;

QY 6 TILDHNPIMKLIIGSFIAQSFESYLYROYOKLSETKLPVLEDEIDETFKSRNY 65
DB 30 tlfsamalpmetvgnfvmlylfetyldlqtlaklpctkvtylsqekfeksray 89
QY 66 SRKAKFSIFGDVYNLAOKLVFIKYDLPKIKWMAVSLNLAVP-VREHAWSTVAQSLCF 124
DB 90 sldkyfnfhvnefvlindsaillffglpwwkms----gavlpjlgldpenellhtlsf 145
QY 125 LGLSLSTLVDLPSTYSYSHFVLEEKFGPNKLTVOIMTDMKSLTLAVAGPILYFL 184
DB 146 lagvmtwsgltdlpslystiesrnfnglnqclwmfltdmktgffsvllgppivaali 205
QY 185 KIEDEPTDFLWYIMVFLVVOILAMTIIPVFIMPMFNKFTPLEDEGLKKSIESIADRVG 244

Db 206 flvkggpylalyaefmflslvmntlypvlaplfnkflrplpdgdlreklekclasslk 265
Qy 245 FPLDKIFVDSGRSSHSAYFTGLPFTSKRIVLEFDITLVNS-NSUDEITAVLAHEIGHMO 303
Db 266 flklklfvdgstrsshshaymgyf-fknktivlydtliqgcknedelayavlahelghnk 324
Qy 304 KHHYNNVFEQLHTFLFLFTSFTYRNTSPYNTFGFLEKSTGSEVDVITKEPPIIG 363
Db 325 lnhctysflavqlaflfgygylvrnstldfrsfqfctg-----pvlfg 369
Qy 364 FMLFDLPLECAMQFVMSLSRTHVQADAVAKKLGKONLCRALIDLOIKNSTMWY 423
Db 370 lllfhtvlpqlhivstfgnlvstrafetqadafavklgyakdlpalvklgeenlsamnt 429
Qy 424 DPLVSSYHSHPTELAERSTALDYSEKKK 452
Db 430 dplysaeyhshpplverlraidd-gecdk 456

RESULT 3
AAG15513
ID AAG15513 standard; Protein: 424 AA.
XX AAG15513;
AC
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 15794.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134370.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.

PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139452.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.

| | | | |
|----|-------------|--------------|--------------|
| PR | 03-AUG-1999 | 99US-0147038 | 99US-0161993 |
| PR | 04-AUG-1999 | 99US-0147204 | 99US-0162142 |
| PR | 04-AUG-1999 | 99US-0147302 | |
| PR | 05-AUG-1999 | 99US-0147192 | |
| PR | 05-AUG-1999 | 99US-0147260 | |
| PR | 06-AUG-1999 | 99US-0147303 | |
| PR | 06-AUG-1999 | 99US-0147416 | |
| PR | 09-AUG-1999 | 99US-0147493 | |
| PR | 09-AUG-1999 | 99US-0147935 | |
| PR | 10-AUG-1999 | 99US-0148171 | |
| PR | 12-AUG-1999 | 99US-0148319 | |
| PR | 13-AUG-1999 | 99US-0148341 | |
| PR | 13-AUG-1999 | 99US-0148565 | |
| PR | 16-AUG-1999 | 99US-0148684 | |
| PR | 16-AUG-1999 | 99US-0149368 | |
| PR | 17-AUG-1999 | 99US-0149175 | |
| PR | 18-AUG-1999 | 99US-0149426 | |
| PR | 20-AUG-1999 | 99US-0149722 | |
| PR | 20-AUG-1999 | 99US-0149723 | |
| PR | 20-AUG-1999 | 99US-0149929 | |
| PR | 23-AUG-1999 | 99US-0149902 | |
| PR | 23-AUG-1999 | 99US-0149930 | |
| PR | 25-AUG-1999 | 99US-0150566 | |
| PR | 26-AUG-1999 | 99US-0150884 | |
| PR | 27-AUG-1999 | 99US-0151066 | |
| PR | 27-AUG-1999 | 99US-0151080 | |
| PR | 30-AUG-1999 | 99US-0151303 | |
| PR | 31-AUG-1999 | 99US-0151438 | |
| PR | 01-SEP-1999 | 99US-0151930 | |
| PR | 07-SEP-1999 | 99US-0152353 | |
| PR | 10-SEP-1999 | 99US-0153070 | |
| PR | 13-SEP-1999 | 99US-0153758 | |
| PR | 15-SEP-1999 | 99US-0154018 | |
| PR | 16-SEP-1999 | 99US-0154039 | |
| PR | 20-SEP-1999 | 99US-0154779 | |
| PR | 22-SEP-1999 | 99US-0155139 | |
| PR | 23-SEP-1999 | 99US-0155486 | |
| PR | 24-SEP-1999 | 99US-0155659 | |
| PR | 28-SEP-1999 | 99US-0156458 | |
| PR | 29-SEP-1999 | 99US-0156596 | |
| PR | 04-OCT-1999 | 99US-0157117 | |
| PR | 05-OCT-1999 | 99US-0157753 | |
| PR | 06-OCT-1999 | 99US-0157865 | |
| PR | 07-OCT-1999 | 99US-0158029 | |
| PR | 08-OCT-1999 | 99US-0158232 | |
| PR | 12-OCT-1999 | 99US-0158369 | |
| PR | 13-OCT-1999 | 99US-0159293 | |
| PR | 13-OCT-1999 | 99US-0159294 | |
| PR | 13-OCT-1999 | 99US-0159295 | |
| PR | 14-OCT-1999 | 99US-0159329 | |
| PR | 14-OCT-1999 | 99US-0159330 | |
| PR | 14-OCT-1999 | 99US-0159331 | |
| PR | 14-OCT-1999 | 99US-0159637 | |
| PR | 14-OCT-1999 | 99US-0159638 | |
| PR | 18-OCT-1999 | 99US-0159584 | |
| PR | 21-OCT-1999 | 99US-0160741 | |
| PR | 21-OCT-1999 | 99US-0160767 | |
| PR | 21-OCT-1999 | 99US-0160768 | |
| PR | 21-OCT-1999 | 99US-0160770 | |
| PR | 21-OCT-1999 | 99US-0160814 | |
| PR | 21-OCT-1999 | 99US-0160815 | |
| PR | 22-OCT-1999 | 99US-0160980 | |
| PR | 22-OCT-1999 | 99US-0160981 | |
| PR | 22-OCT-1999 | 99US-0160989 | |
| PR | 25-OCT-1999 | 99US-0161404 | |
| PR | 25-OCT-1999 | 99US-0161405 | |
| PR | 25-OCT-1999 | 99US-0161406 | |
| PR | 26-OCT-1999 | 99US-0161359 | |
| PR | 26-OCT-1999 | 99US-0161360 | |
| PR | 26-OCT-1999 | 99US-0161361 | |
| PR | 28-OCT-1999 | 99US-0161920 | |
| PR | 28-OCT-1999 | 99US-0161992 | |
| PR | 28-OCT-1999 | 99US-0161993 | |
| PR | 29-OCT-1999 | 99US-0162142 | |
| PR | 29-OCT-1999 | 99US-0162143 | |
| PR | 29-OCT-1999 | 99US-0162144 | |
| PR | 29-OCT-1999 | 99US-0162145 | |
| PR | 29-OCT-1999 | 99US-0162146 | |
| PR | 29-OCT-1999 | 99US-0162147 | |
| PR | 29-OCT-1999 | 99US-0162148 | |
| PR | 29-OCT-1999 | 99US-0162149 | |
| PR | 29-OCT-1999 | 99US-0162150 | |
| PR | 29-OCT-1999 | 99US-0162151 | |
| PR | 29-OCT-1999 | 99US-0162152 | |
| PR | 29-OCT-1999 | 99US-0162153 | |
| PR | 29-OCT-1999 | 99US-0162154 | |
| PR | 29-OCT-1999 | 99US-0162155 | |
| PR | 29-OCT-1999 | 99US-0162156 | |
| PR | 29-OCT-19 | | |

PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132863.
PR 14-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142055.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 16-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144684.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 05-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 06-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 13-AUG-1999; 99US-0149368.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.

| | | | | |
|-----------------------|------------------|--------------------|------------|-------------|
| Query Match | 36.2%; | Score 848; | DB 21; | Length 419; |
| Best Local Similarity | 41.7%; | Pred. NO. 1.1e-81; | | |
| Matches 182; | Conservative 71; | Mismatches 159; | Indels 24; | Gaps 6 |

| | | |
|----------|--------|--|
| | RESULT | 5 |
| AAV26896 | ID | AAV26896 standard; Protein; 475 AA. |
| XX | AC | AAV26896; |
| XX | DT | 07-DEC-1999 (first entry) |
| XX | DE | Human farnesylated--protein converting enzyme 1 protein. |
| KM | XX | Human; farnesylated-protein converting enzyme; Face; homology; Afc-1 |
| WW | XX | Saccharomyces cerevisiae; proteolysis; prenylation; metalloprotease; |

| | | | | |
|-----------------------|------------------|--------------------|------------|-------------|
| Query Match | 29.88; | Score 698.5; | DB 20; | Length 475; |
| Best Local Similarity | 35.68; | Pred. No. 1.2e-65; | | |
| Matches 165; | Conservative 75; | Mismatches 163; | Indels 61; | Gaps 7 |

```

QY      22 FSIQOFSEFSLTYVRQ---YOKISETKYLPVLDEIDEDFFHHKSRNRSRAKAFSEJGDV 78
Dp      26 fswtvywlwelflaqrtriyk--ttlvhpelqgimdselctsfirlyqldksctsfwgl 83
QY      79 YNLQKLVFTKYDLEFPKIMMAVSLNAVLVPVRHNVSTAQSLCEJGLSLSTLVDLP 138
Dp      84 ysetegrlilllfgiplyrlwlsgrfcgy---agfgpeyltqslvlllatltsaltgipr 140
QY      139 LSYSHFVLEBEKGFENKLTVQOLWITDMIKSLTLAVAGPILYFLKIDFKPTDLWYI 198
Dp      141 wslyntfvlceekhgfnqctlgfemkdaikkfvtvqcillrvssalllyllykigdcyffiya 200
QY      199 MVEFLVVOILLMTIIPFIMPMFKFTPLDEDELKAKIESLADRVGPPDLKIVYDQSKR 258
Dp      201 wlfvlsvlavlvtlyadylapipldtkftrlppegkllkeelwemaksldlplkvyvvvegskr 260
QY      259 SSHNAVETGQPFPSKRTVLEPDTLVNSNS-----287
Dp      261 sshsnayfyif-fknktivlfdtllleeyvslnkdigedsmeprneeegnseetkavxn 319
QY      288 -----DELTAVLAHEIGHMOKNHNINMYVINSOLAHFLJFSLFTSIYRNTSEYNTGFF 341
Dp      320 kkggckheevlavyghelgmklyghtvknllisgmssmfcfflfaavgkelfaaafgy 379
QY      342 LEKSTGSEVDPVITKEPPIIGFM-LFNDDLTPPLECAMQFVMSLISHTHEYOADAAYAKUL 400
Dp      380 dsq-----puliqlllltqfflspnevlstclvlarrrfctfgdatakl 425
QY      401 GYKONLCRALDIQIKKLSTWNVNDPLVSSYHYSHTPLAESSTVL 444

```

Db 426 gkakdlysaliklnkdnlgfpvswlfsmwhysbpllerlqal 469

RESULT 6

AAW87587
ID AAW87587 standard; Protein: 474 AA.

AC AAW87587;

DT 12-MAR-1999 (first entry)

DE Human metalloprotease AFCL.

XX human AFCL metalloprotease; AFCL; haFCL; vaccine; cancer; inflammation;
KW autoimmunity; allergy; asthma; rheumatoid arthritis; osteoarthritis;
KW neurological abnormality; septic shock; sepsis; stroke; osteoporosis;
KW ischemia reperfusion injury; cardiovascular disease; kidney disease;
KW liver disease; myocardial infarction; hypertension; AIDS; hypotension;
KW haematological abnormality; aplastic anaemia; male pattern baldness;
infection.

XX Homo sapiens.

PN EP887416-A2.

PD 30-DEC-1998.

PE 18-JUN-1998; 98EP-0304827.

PR 12-FEB-1998; 98US-0022699.

PR 24-JUN-1997; 97EP-0304440.

PA (SMIK) SMITHKLINE BEECHAM CORP.

PA (SMIK) SMITHKLINE BEECHAM PLC.

PI Kikly K, Knab A, Southan C;

DR WPI: 1999-047883/05.

DR N-PSDB: AAV83752.

PT New human metalloprotease (haFCL) polypeptides and polynucleotides -

PT useful as diagnostic reagents and for prevention and treatment of

PT cancer, AIDS and stroke

PS Claim 4: Page 14; 22pp; English.

XX The present sequence represents a human AFCL metalloprotease (haFCL)
CC polypeptide. haFCL polypeptides and polynucleotides are useful for
CC diagnosing diseases related to over or underexpression of haFCL protein
CC by identifying mutations in the haFCL gene, and/or analysing for the
CC presence or amount of expressed polypeptide. haFCL polypeptides can be
CC used to screen for agonists and antagonists by measuring the binding to
CC polypeptide, and observing the stimulation or inhibition of polypeptide
CC function. haFCL polypeptides can be administered directly or as a
CC vaccine to inoculate against disease. Diseases diagnosed, prevented and
CC treated include: cancer, inflammation, autoimmunity, allergy, asthma,
CC rheumatoid arthritis, neurological abnormalities, septic shock, sepsis,
CC stroke, osteoporosis, osteoarthritis, ischemia reperfusion injury,
CC cardiovascular disease, kidney and liver disease, myocardial infarction,
CC hypo/hypertension, AIDS, haematological abnormalities, aplastic anaemia,
CC male pattern baldness, and bacterial, fungal, protozoan and viral
CC infections. The haFCL polypeptide is also useful for mapping the gene to
CC a chromosome, allowing gene inheritance to be studied through linkage
CC analysis.

XX Sequence 474 AA;

Query Match 27.9%; Score 653; DB 20: Length 474;

Best Local Similarity 34.7%; Pred. No. 8.3e-61;

Matches 161; Conservative 73; Mismatches 168; Indels 62; Gaps 8;

QY 22 FSIADFESYLYTRQ--YOKLSETKLPVLEDEIDETFRHNSRAKAFSIFGDV 78

Db 26 fswtvytwetflaqrgrlyk--ctthvppeqlgmdsetfekrlyqldstcfswgl 83

QY 79 YNLAOKLVFIKYDLPFKKIMHMAVSLNANLVPVRFHMVSTVAOSLCFGLSSSLTVLDP 138

Db 84 ysetegtlilllfggipyrlwrlsgrfcgy--agfgeyelqstlwflllatlfsalglp 140

QY 139 LSYSHFVLEKEFGFNKLTVOIMIDMKSLTLAVALGCPILYLFKLFDEKFPDPLMYI 198

Db 141 wslntvlyeehngfngqlgffmdaalkvtvtqcllpvsslllylkigdyffilya 200

QY 199 WVEFLFVQILAMTIIPVFMPEFNKFTLEDGELKKSIESLADRVGFPLDKITFYDGSKR 258

Db 201 wlfetvslvltlyadylapfdkftlpqgklleevmaksdldlktkvyvegskr 260

QY 259 SSHSNAYFTGLPFPSKRIVLFDTLVNSNS----- 287

Db 261 sshsnayfygffkxkrlvlfcllleesvlnkddqedsqmeprneegnsaelkavkn 319

QY 288 -----TDEITAVLAHEIGHQKNHIVMNVIFSOHTFLISLFTSIRNHSFYTFGGF 341

Db 320 kkgckneevlavlgheghw-kwdlqskislarefpvflfavlgrkellaaigly 378

QY 342 LEKSTGSEFVDVITKEPFIIGFM-LFNDLTPLECAMQFVMSLSTRHEXQADAVAKKL 400

Db 379 dsq-----pfligllliffgflspynevlsfclvtlsrrfegadefakl 424

QY 401 GYKONLCRALDLOITKNTSTMNVDPLVSSYHSHTLAERSTAL 444

Db 425 gkakdlysaliklnkdnlgfpvswlfsmwhysbpllerlqal 468

RESULT 7

AAW8297
ID AAW8297 standard; Protein: 340 AA.

AC AAW8297;

DT 14-MAR-2001 (first entry)

DE Lung cancer associated polypeptide sequence SEQ ID 635.

XX Human; lung cancer associated protein; neuroprotective; cytostatic;

XX cardioactive; immunomodulatory; muscular active; vulnereary;

XX gastrointestinal; nephrotropic; antinfetive; gynecological;

XX antibacterial; diagnosis; neural disorder; immune disorder; reproductive;

XX proliferative disorder; wound healing; infectious disease.

OS Homo sapiens.

PN WO20005180-A2.

XX 21-SEP-2000.

PE 08-MAR-2000; 2000WO-US05918.

PR 12-MAR-1999; 99US-0124270.

PA (HUMA-) HUMAN GENOME SCI INC.

PA (ROSE/) ROSEN C A.

PI Ruben SM;

XX WPI: 2000-587514/55.

DR N-PSDB: AAF18173.

PT Lung cancer associated gene sequences, referred to as lung cancer

PT antigens, useful for treatment, prevention, and diagnosis of disorders

PT such as lung cancer -

PS Claim 11: Page 1132-1134; 1425pp; English.

CC Polynucleotide sequences AAF17982 - AAF18424 encode human lung cancer


```

RESULT 9
AAB46348
ID AAB46348 standard; Protein: 414 AA.
XX
AC AAB46348:
XX
DT 05-APR-2001 (first entry)
XX
DE H. pylori HPS187 protein.
XX
KM Microbial infection; antibacterial; Helicobacter pylori infection;
KM vaccine; screening.
XX
OS Helicobacter pylori.
XX
PN WO200073502-A2.
XX
PD 07-DEC-2000.
XX
PF 31-MAY-2000; 2000MO-EP05024.
XX
PR 31-MAY-1999; 99DE-1024965.
PR 17-JUN-1999; 99DE-1027740.
PR 21-JUL-1999; 99DE-1034029.
XX
PA (PLAC ) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
PA (CREA-) CREATOGEN GMBH.
XX
PI Apfel H, Fuchs TM, Glibbs CP, Hueck CJ, Meyer TF;
XX
DR WPI: 2001-049948/06.
DR N-PSDB; AAF25625.
XX
PT Preparing an agent for diagnosis or control of microbial infection,
PT useful particularly against Helicobacter, based on identification of
PT essential genes in defective mutants -
XX
PS Claim 37; Page 356-357; 366pp; German.
XX
CC This invention describes a novel preparation of an agent (A) for
CC detection, prevention and/or treatment of microbial infection by:
CC (i) identifying essential genes (I) and corresponding polypeptides
CC (II); (ii) identifying compounds that are directed against (II) and
CC inactivate the microbe; (iii) testing these for suitability for use; and
CC (iv) formulating selected (A). Identifying essential genes (I) comprises
CC preparation of gene-deficient microorganisms by conditional antisense
CC inhibition (CAI) and/or subtractive recombination mutagenesis (SRM),
CC then determining viability and/or survival of the deficient organisms.
CC The products of the invention have antibacterial activity. (A) (which may
CC be a nucleic acid (ia), vector or host cell containing (ia), derived
CC polypeptide (Iia), or fragments, (Iia)-specific antibodies or their
CC fragments or an inhibitor of (Iia)) are particularly used for diagnosis,
CC treatment or prevention of infection by Helicobacter pylori. Particularly
CC (Iia) and (Iia) are used in DNA, subunit or live vaccines. The method
CC identifies essential genes, including those that have homologs in other
CC species, so identified (A) should have a broad spectrum of activity. Many
CC gene-deficient cells can be screened quickly, in an automated process.
CC and the identified genes can be used for screening without purification.
XX
SO Sequence 414 AA:

```

Query Match 19.4%; Score 454; DB 22; Length 414;
 Best Local Similarity 30.7%; Pred. No. 1.2e-39;
 Matches 129; Conservative 84; Mismatches 151; Indels 56; Gaps 17;

```

OY 39 OKLSTKLPVLEIDIDETFRKSRNYSRAKAFSIFGVDVYLAOKLVFIKYDLPPKIMH 98
DB 42 qlyce---kpyl---lpqkdyeeagnyalrkmqlslisqld---gllfagv-vffgltlh 91
OY 99 M-AVSLNANVLPRFHHVSTVAVOGLCLGLSLISSTLVDPDPLSYSHFVLEERKGFNKL 156
DB 92 ledlthyn-lp---etlgyivfallfnaigsyia---lpisyttcmhdikelfgfskv 142

```

```

OY 157 TWQLWITDMIKSLTTLAVAGCPILYLFLKIPDKFPPTDFLWYIMF--LFVQIILAMTIIP 214
DB 143 slsllfkdfkglstlslsvgllllyclimlthveh---welslffvfvfmlanllyfp 199
OY 215 VFIIMPFNKFPLEDEGELKKSIESLADRYGFLDKIFVIDGSRSSHNAVPTGLPFTSK 274
DB 200 k-laglfngfuplnnrldiesqlegmmdkvgfksqglfvmddskrtgrlnayfglg-knk 257
OY 275 RLVFDFTLVNSSTDEITAVLAHEIGHWOKNHIYVMVIFSQLHFLIFSLETS---IYR 330
DB 258 rvlldftliskvgtegliallghelghfknkollksjgimgliallallahlpplvfe 317
OY 331 NTSFYNTFGEFLEKSTGSPVDPIVIRKEPFIIGEM-LTNDLLTPELCAMQFMSLSRTH 389
DB 318 -----gfnvsglpsall-----aillflfpvtsfyampI-----lgtfsfkn 354
OY 390 EXQADAVAKKLIGYKONLCRALIDLOIKNLSTWNVPLSSYSHVSHPTLAERSTALDYSE 449
DB 355 eynackfgaslskskevlakalvsivsenkafpysphfyfhlftphlperlkaldyle 414

```

RESULT 10
 AAB95578
 ID AAB95578 standard; Protein: 235 AA.
 XX
 AC AAB95578:
 XX
 DT 26-JUN-2001 (first entry)
 XX
 DE Human protein sequence SEQ ID NO:18234.
 XX
 KW Human; primer; detection; diagnosis; antisense therapy; gene therapy.
 OS Homo sapiens.
 XX
 PN EP1074617-A2.
 XX
 PD 07-FEB-2001.
 XX
 PF 28-JUL-2000; 2000EP-0116126.
 XX
 PR 29-JUL-1999; 99JP-0248036.
 PR 27-AUG-1999; 99JP-0300253.
 PR 11-JAN-2000; 2000JP-0118776.
 PR 02-MAY-2000; 2000JP-0183767.
 PR 09-JUN-2000; 2000JP-0241899.
 XX
 PA (HELI-) HELIX RES INST.
 XX
 PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
 XX
 DR WPI: 2001-318749/34.
 XX
 PT Primer sets for synthesizing polynucleotides, particularly the 5602
 PT full-length cDNAs defined in the specification, and for the detection
 PT and/or diagnosis of the abnormality of the proteins encoded by the
 PT full-length cDNAs -
 XX
 PS Claim 8; SEQ ID 18234; 2537pp + CD ROM; English.
 XX
 CC The present invention describes primer sets for synthesizing 5602
 CC full-length cDNAs defined in the specification. Where a primer set
 CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
 CC to the complementary strand of a polynucleotide which comprises one of
 CC the 5602 nucleotide sequences defined in the specification, where the
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
 CC of an oligonucleotide comprising a sequence complementary to the
 CC complementary strand of a polynucleotide which comprises a 5'-end
 CC sequence and an oligonucleotide comprising a sequence complementary to a
 CC polynucleotide which comprises a 3'-end sequence, where the
 CC oligonucleotide comprises at least 15 nucleotides and the combination of


```

Db      53 vaegsnvknflffigvpldwflffvllvsgvskkikkwleavpfrf-----lqtvqf 106
QY      125 LGLSLSLVLDPLSLSYSHFLLEKFGFNKLTVOIMITDMKSLTLAATG-----PILY 181
Db      107 VVLSILTLVLPLDWDYGYSLD-YNLSQTLSWAKDQYISFISWISPIITLCLVIVY 165
QY      182 LFLKLFDFKFPDFLWYIMW----FLFVVOILAMTIIIPVFMFMFKFPLEDEGLKKSIE 237
Db      166 wlkthekkwwlyawllcvrfsllffifq-----pvlidplyndfplrkkeleskl 218
QY      238 SLADRVGPEPLDKIEVIDGSRSSHSNAYFTGLPFTSKRIVLPDVLVNSNSTDEITAVLAH 297
Db      219 eladeanlpdphvyemsektnalnayvtgig-ankrtivldtclnklddseillfimg 277
QY      298 ELGHQOKNHIYVMVIFESQHTFLIFSL-----FTSIRNTSYFNFGFLEKSTGSFVDP 352
Db      278 emghyvmkh-----vyiglagyllvslagfyvdiklyktrvlrslsmfhlegrhdaalp 332
QY      353 VITKEFPIITIGFMLENDLITPLECAMQFVMSLSRTHETQADAYAKKL-GYKQNLRALI 411
Db      333 l-----llllsvlsfavtfsna-----vsrygenkadqylteltenreaavkftq 379
QY      412 DLQIKNLSTMNVDPLYSSTHYSHPTLAERSTALDYVESEKKN 453
Db      380 dlavtqlsqvdpvlyvkllfrgshpsimeriqh-----aekeen 417

RESULT 14
AAW20512 standard; Protein; 192 AA.
XX
AC      AAW20512;
XX
DT      14-JUL-1997 (first entry)
XX
DE      H. pylori cell envelope protein 4714375.aa.
XX
KW      Cytoplasmic; vaccine; prevention; treatment; infection; identification;
KW      binding compound; bacterium; life cycle; activator; bacteria; inhibitor;
KW      duodenal ulcer disease; chronic gastritis; diagnosis; envelope.
XX
OS      Helicobacter pylori.
XX
FH      Key Location/Qualifiers
FT      Misc-difference 51 /label= unknown
FT      /note= "encoded by SGC"
XX
PN      WO9640893-A1.
XX
PD      19-DEC-1996.
XX
PF      06-JUN-1996; 96WO-US09122.
XX
PR      01-APR-1996; 96US-0630405.
XX      07-JUN-1995; 95US-0487032.
XX
PA      (ASTR ) ASTRA AB.
XX
PI      Berglindh OT, Smith D, Mellgaerd BL;
XX
DR      WPI: 1997-052306/05.
XX      N-PSDB: AAT67669.
XX
PT      Helicobacter pylori nucleic acid sequences and related
PT      polypeptide(s) - useful for vaccines to treat or prevent H. pylori
PT      infection, and to detect Helicobacter
XX
PS      Claim 56; Page 673; 1481pp; English.
XX
CC      This sequence is a H. pylori cell envelope protein.
CC      The protein may be used in a vaccine to prevent or treat H. pylori
CC      infection or to identify H. pylori polypeptide binding compounds,

```

```

CC      useful as potential H. pylori life cycle activators or inhibitors.
CC      The genomic sequence of H. pylori (ATCC 55679) was determined from
CC      overlapping contigs generated by mechanically shearing the bacterial
CC      DNA. The sequences were analysed for ORF of at least 180 nucleotides,
CC      and the predicted coding regions defined by computer evaluation. No
CC      identity likely H. pylori antigens for vaccine development, the amino
CC      acid sequences predicted from various ORF were analysed for significant
CC      homology to other known or exported membrane proteins. Having identified
CC      and determined the sequences of interest, particular regions can be
CC      isolated from H. pylori by PCR amplification for recombinant polypeptide
CC      production, e.g. in E. coli hosts.
XX
SQ      Sequence 192 AA;
XX
Query Match 11.7%; Score 274.5; DB 18; Length 192;
Best Local Similarity 33.2%; Pred. No. 5.2e-21;
Matches 72; Conservative 40; Mismatches 74; Indels 31; Gaps 8;
QY      239 LADRVGPEPLDKIEVIDGSRSSHSNAYFTGLPFTSKRIVLPDVLVNSNSTDEITAVLAHE 298
Db      1 mmokvqfksqglfmdaskrdgrlnaytggig-knkrrvllfcllsvkgtexllallghe 59
QY      299 IGHWO-KNHIYVMVIFESQHTFLIFSLFTS-----IYRNTSFYNTGCFLEKSTGSFVDPV 353
Db      60 lghfkxkdlknlglimggl-lalvfallahlpvlfe-----gfnvsgpsall---- 107
QY      354 ITKEFPIITIGFMLENDLITPLECAMQFVMSLSRTHETQADAYAKKLGYKQNLRALID 412
Db      108 -----llllflpvtfsfyampI-----lghfskneyackfgaslskctakalvs 155
QY      413 DLQIKNLSTMNVDPLYSSTHYSHPTLAERSTALDYVE 449
Db      156 lvenkafpshpbyvflhftnpplerlkaldyele 192

RESULT 15
AAV05502 standard; Protein; 211 AA.
XX
AC      AAV05502;
XX
DT      05-JUL-1999 (first entry)
XX
DE      Bacillus subtilis metalloprotease YHFN polypeptide (aa198-426).
XX
KW      Metalloprotease; protease; YHFN; textile; detergent; feedstuff;
KW      animal feed; host expression system.
XX
OS      Bacillus subtilis.
XX
FH      Key Location/Qualifiers
FT      Region 80..84 /note= "HEXXH metalloprotease motif"
XX
PN      WO9914341-A2.
XX
PD      25-MAR-1999.
XX
PF      08-SEP-1998; 98WO-US18677.
XX
PR      15-SEP-1997; 97GB-0019637.
XX
PA      (GENV ) GENENCOR INT BV.
XX      (GENV ) GENENCOR INT INC.
XX
PI      Estell DA;
XX
DR      WPI: 1999-244036/20.
XX
XX      New metalloproteases derived from gram-positive microorganisms
XX
CC      Claim 15; Fig 3; 54pp; English.

```

xx This polypeptide corresponds to amino acid residues 198-426 of
CC YHFN (see AAY05501), a novel metalloprotease (MP) of the
CC Gram-positive microorganisms *Bacillus subtilis* strain I-168.
CC The MP polypeptide shows homology to *Bacillus stearothermophilus*
CC subtilisin J. The invention relates to the identification of
CC novel MPs in Gram-positive microorganisms, especially *B. subtilis*,
CC *B. stearothermophilus*, *Bacillus licheniformis* and *Bacillus*
CC *amyloliquefaciens*. The invention also provides methods for the
CC production of MPs in host cells as well as for the production of
CC heterologous proteins in a host cell in which all or part of the MP
CC gene is mutated or deleted. Gram positive MPs have applications in
CC the textile industry, in cleaning compositions (claimed), and in
CC animal feed. Claimed cleaning compositions can include the present
CC polypeptide.
xx

SQ Sequence 211 AA:

Query Match 9.6%; Score 225.5; DB 20; Length 211;
Best Local Similarity 28.0%; Pred. No. 1e-15;
Matches 68; Conservative 43; Mismatches 85; Indels 47; Gaps 8;

QY 217 IMPMENKFPLEDEGELKKSIESLADRVGFPLDKIEVIDGSKRSSHSMAYFTGLPTSKRI 276
Db 1 idplyndfyplknkelskleleadeanipadhvevmsektnainayvtgig-ankri 59
QY 277 VLPDITLVNSNSDEITAVLAHEIGHQKNHIVNMFISQLHTFLIFSL-----FTSIRN 331
Db 60 vldtclnklddsellfmgmhgmhkh-----vylglaagllvslagfyvldklykr 114
QY 332 TSFYNTFGFLEKSTGSEFVDPVITKEFPITIGFMLEFNDLTPLECAMQFVMSLISRTHEY 391
Db 115 tvrltsmfhleg-----hd--tpfsna-----vstrygen 143
QY 392 QADATAKKL-GYKQNLKRALIDQIKNLSTMNVDPLYSSYHYSHPTLAERSTALDYVSEK 450
Db 144 kadqygleltenreaavkftfgdlavtqlsqvdpvlykifrgshpsimeriqh---aek 199
QY 451 KKN 453
Db 200 een 202

Search completed: February 12, 2002, 09:27:24
Job time: 110 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 12, 2002, 09:26:13 ; Search time 43.94 Seconds

(without alignments)
1507.996 Million cell updates/sec

Title: US-09-165-460A-2
Perfect score: 2342
Sequence: 1 MEDLTIIDHPNIPWKLIIIS.....HPTLAERSTALDYSEKKKN 453

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :
1: SPREMBL_17:*
2: sp_archaea:*
3: sp_bacteria:*
4: sp_fungi:*
5: sp_human:*
6: sp_invertebrate:*
7: sp_mammal:*
8: sp_mhc:*
9: sp_organelle:*
10: sp_phage:*
11: sp_plant:*
12: sp_protist:*
13: sp_virus:*
14: sp_vertebrate:*
15: sp_unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|--------|-------------|--------|----|-------------|
| 1 | 1085.5 | 46.3 | 462 | 3 | 09C2C0 |
| 2 | 819.5 | 35.0 | 459 | 10 | 09M139 |
| 3 | 730.5 | 31.2 | 451 | 5 | 09V7M7 |
| 4 | 654 | 27.9 | 316 | 10 | 004602 |
| 5 | 612.5 | 26.2 | 442 | 5 | 09XVES |
| 6 | 501.5 | 19.6 | 395 | 2 | 09P14 |
| 7 | 459 | 19.4 | 407 | 2 | 09ZKE0 |
| 8 | 454 | 19.4 | 407 | 2 | 025144 |
| 9 | 406.5 | 17.4 | 447 | 5 | 09V7M6 |
| 10 | 378 | 16.1 | 428 | 1 | 09HSA2 |
| 11 | 372.5 | 15.9 | 426 | 5 | 09V7M5 |
| 12 | 241.5 | 10.3 | 504 | 5 | 09V7H6 |
| 13 | 197 | 8.4 | 338 | 1 | 09V0G6 |
| 14 | 152 | 6.5 | 291 | 2 | 09I0I3 |
| 15 | 148.5 | 6.3 | 279 | 2 | 09JVI9 |
| 16 | 148.5 | 6.3 | 347 | 2 | 09K006 |
| 17 | 136 | 5.8 | 287 | 2 | 09KST9 |
| 18 | 135.5 | 5.8 | 289 | 2 | 09PA93 |
| 19 | 132.5 | 5.7 | 760 | 1 | 09UX10 |

| | | | | | | |
|----|-------|-----|------|----|--------|--------------------|
| 20 | 129.5 | 5.5 | 304 | 2 | 09K9B6 | 09K966 bacillus ha |
| 21 | 128 | 5.5 | 2298 | 8 | 09B1K6 | 09B1K6 lotus japon |
| 22 | 125 | 5.3 | 316 | 2 | 09A5E1 | 09A5E1 caulobacter |
| 23 | 123 | 5.3 | 315 | 1 | 029337 | 029337 archaeoglob |
| 24 | 122 | 5.2 | 3119 | 5 | 025857 | 025857 plasmidum |
| 25 | 120.5 | 5.1 | 289 | 1 | 09U2K3 | 09U2K3 pyrococcus |
| 26 | 120.5 | 5.0 | 1436 | 10 | 09A1D5 | 09A1D5 streptococ |
| 27 | 118 | 5.0 | 324 | 2 | 09A1D5 | 09A1D5 guillardia |
| 28 | 117.5 | 5.0 | 324 | 2 | 09C1I2 | 09C1I2 lactococcus |
| 29 | 115.5 | 4.9 | 485 | 1 | 09H1C1 | 09H1C1 thermoplas |
| 30 | 112.5 | 4.8 | 638 | 5 | 002260 | 002260 caenorhabd |
| 31 | 110.5 | 4.7 | 355 | 5 | 017519 | 017519 caenorhabd |
| 32 | 110.5 | 4.7 | 2280 | 8 | 09MER2 | 09MER2 oenothera h |
| 33 | 109.5 | 4.7 | 669 | 2 | P74499 | P74499 synecocyst |
| 34 | 108.5 | 4.6 | 913 | 5 | 097272 | 097272 synecocyst |
| 35 | 107.5 | 4.6 | 236 | 8 | 09T241 | 09T241 phytophthor |
| 36 | 106.5 | 4.5 | 521 | 13 | 09Y164 | 09Y164 fundulus he |
| 37 | 106.5 | 4.5 | 645 | 12 | 092491 | 092491 bombyx mori |
| 38 | 105.5 | 4.5 | 842 | 10 | 049423 | 049423 arabidopsis |
| 39 | 105.5 | 4.5 | 303 | 2 | P96088 | P96088 thermoplas |
| 40 | 105.5 | 4.5 | 478 | 2 | 09ZV24 | 09ZV24 staphylococ |
| 41 | 105 | 4.5 | 572 | 8 | 099X20 | 099X20 staphylococ |
| 42 | 105 | 4.5 | 1191 | 5 | 09V0T8 | 09V0T8 drosophila |
| 43 | 104.5 | 4.5 | 350 | 2 | 030362 | 030362 thermocae |
| 44 | 104.5 | 4.5 | 499 | 2 | 09CG49 | 09CG49 lactococcus |
| 45 | 104 | 4.4 | 333 | 11 | 09JKF0 | 09JKF0 rattus norv |

ALIGNMENTS

| RESULT | ID | PRELIMINARY: | PRT: | 462 AA. |
|-----------------------|--|--|------|---------|
| 09C2C0 | 09C2C0 | | | |
| AC | 09C2C0 | | | |
| DT | 01-JUN-2001 (TREMBL) | 17, Created) | | |
| DT | 01-JUN-2001 (TREMBL) | 17, Last sequence update) | | |
| DT | 01-JUN-2001 (TREMBL) | 17, Last annotation update) | | |
| DE | PROBABLE ZINC METALLO-PROTEASE. | | | |
| GN | B1N2.80. | | | |
| OS | Neurospora crassa. | | | |
| OC | Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes. | | | |
| OC | Sordariales; Sordariaceae; Neurospora. | | | |
| OX | NCBI_TaxID=5141; | | | |
| RN | [1] | | | |
| RP | SEQUENCE FROM N.A. | | | |
| RA | Schulte U., Algen V., Hehnel J., Brandt P., Farlmann B., Holland R., | | | |
| RA | Nyakatura G., Mewes H.W., Mannhaupt G., | | | |
| RL | Submitted (FEB-2001) to the EMBL/Genbank/DBJ databases. | | | |
| RN | [2] | | | |
| RP | SEQUENCE FROM N.A. | | | |
| RA | German Neurospora genome project. | | | |
| RL | Submitted (FEB-2001) to the EMBL/Genbank/DBJ databases. | | | |
| DR | EMBL; AL513444; CAC28689.1; | | | |
| KW | Protease. | | | |
| SO | SEQUENCE 462 AA; 52825 MW; 6B4ED30AC028EC CRC64; | | | |
| Query Match | 46.3%; Score 1085.5; DB 3; Length 462; | | | |
| Best Local Similarity | 47.3%; Pred. No. 3.2e-69; | | | |
| Matches | 212; Conservative 79; Mismatches 140; Indels 17; Gaps 3; | | | |
| Db | 4 | LKTIDHPNIPWKLIIISFSTIAQSFESYLYTRQOKSERKLPVLEDEIDDEFHRSR | 63 | |
| Db | 7 | LARFLDRPLFPWKLLIISFSLANFYIEGFLGRQVLRKTRKVPVLEHEVSQEVDSQ | 66 | |
| Db | 64 | NYSRAKAFSIFGDVYNLAOKLFTKYDLPFRIMHMAVSLNAVLPVFHMAVSTAQSLC | 123 | |
| Db | 67 | AYGRKAKFEFFSGIYGIONILFFQPDVLRKMSFAGNLVFRAPAFS--GELSGSIV | 124 | |
| Db | 124 | FLGLISLSTLVDPPLSYSHFVLEKEGFKNLTVQVMTIMKISLTAIVAIGPILYLF | 183 | |
| Db | 125 | FVLSFVVISQILSPTSLVHTFVLEKEGFKNSAKLWTKIKISLFLVLTPIPLAGF | 184 | |

184 LKIPDKPTDFLWYIMVFLVVOILAMTIIPIVIMPMNKFPLEDELKKSIESLARV 243
185 LAIVKOTGNOFFYIMVAVAGIQQVMTIITPFIPLPKNKSLPELEGELKSSVEDLAKL 244
244 GEPLOKIFVIDGSKRSSHNAVFTGLPPTSKRIYVLDLVNSNSTDEITAVLAHEIGHMO 303
245 KPELSELHVIDGSKRSSHNAVFTGLPPTSKRIYVLDLVNSNSTDEITAVLAHEIGHMO 303
304 KNIHYNMVIFSOHLTFLFSFTSTYRTSTYNTGFFLEKSTGSPVDPVITKEPIIIG 363
304 LGHTTSLGIGIOAHFAFLFSFVSFINNNSLYADGFGH-----TVHPITVG 349
364 FMEFNDLTPLFCAMQFVMSLISRTHEYOADAAYAKKGYKONLCALDLOIKNSTMV 423
350 FLPSFDVIGPTDYLKIKLGMNVLSKFEFQADFAFKLGYNALNLSLKIQLIQLNSTDA 409
424 DPLVSSYHSHPTLAERSTALDYSEKK 451
410 DMFATYHFSHPILTERLKALNMOSTGK 437

RESULT 2
ID Q9M139 PRELIMINARY: PRT: 459 AA.
AC Q9M139;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE PUTATIVE CAAX PRENYL PROTEASE.
GN A7G01320.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; Core eudicotyledons; Rosidae;
OC Eucosids; II: Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Lamar B., Stoneking T., Stumpf J., Mewes H.W., Lemcke K.,
RA Mayer K.F.X.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AL161491; CAB80941.1;
DR InterPro: IPR001915; Peptidase_M48.
DR Pfam: PF01435; Peptidase_M48; 1.
KW Protease.
SQ SEQUENCE 459 AA; 52502 MW; F7EF6C32DE55E79 CRC64;

Query Match 35.0%; Score 819.5; DB: 10; Length 459;
Best Local Similarity 38.1%; Pred. No. 2.3e-50;
Matches 184; Conservative 73; Mismatches 155; Indels 71; Gaps 9;
QY 13 IPMKLIISGFISIAQFSFYLYROYKLSKTLPPVLEDEIDETPFKSNYSARAKAF 72
DB 3 IPFMETVVGFMIVMVIIEFTYIDLKQLTALKPLPKTLVGIYSQGEFKSNAYR----- 56
QY 73 SIFSDVYNLAQKLVFK-----YDLPRKIMHMAVSLNLAFLP-VRFHMSVTV 118
DB 57 DITENFNIISYGFHFVHFVFTILMDSALIFGILPMFKMS-----GAVLPRLGLDPENEI 112
QY 119 AQSICFGLSSLSLTVLPLSYSHVLEKFGFNKLTVOVMITDMKSLTAVAGTGP 178
DB 113 LHTLSFLAGVMTWSQIDLPFSLVTFVIESRHGKNQOTIMFTRDMIKGTFLSVILGPP 172
QY 179 ILVFLKIFDFPTDFLWYIMVFLVVOILAMTIIPIVIMPMNKFPLEDELKKSIES 238
DB 173 IVAAIIFVQKGGPLATIIYLAHMFILSLVMITIPVIALPKNKFTPLPGDLEKLEK 232
QY 239 LADRVGFPDLKIFVIDGSKRSSHNAVFTGLPPTSKRIYVLDLVNS-NSTDEITAVLAH 297

DB 233 LASSLKFPFLKFLVVDGSTRSSHNAVYMGF-FKKKRRIYVLDTLQOQCKNEDEIYAVLAH 291
QY 298 EIGHQKKNHYNMVIFSOHLTFLFSFTSTYRTSTYNTGFFLEKSTGSPVDPVITKE 357
DB 292 ELGHMKLHNTTYSFIAVOILAFLOFGGYTLVNSNSTDEITAVLAHEIGHMO 337
QY 358 FPIIGFPLFNDLTPLFCAMQFVMSLISRTHEYOADAAYAKKGYKONLCALDLOIK- 416
DB 358 -PVLIGLITFQHTVPLQHLVSPGLNLSKRAFEQADFAFKLGYNALNLSLKIQLIQLNSTDA 396
QY 417 -----NLSTMNVDPLVSSYHSHPTLAERSTALDYSE 449
DB 397 DNNRTQVTSICVTHLNGFVGIQDEENLSAMNPDLSAHYHSHPTLAERSTALDYSE 454
QY 450 KKK 452
DB 455 DKK 457

RESULT 3
ID Q9V7W7 PRELIMINARY: PRT: 451 AA.
AC Q9V7W7;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE CG3000 PROTEIN.
GN CG3000.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BERKELEY;
RX MEDLINE-20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abriil J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Baller R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferriz C., Ferriera S., Fleischmann W.,
RA Foster C., Gabriellian A.E., Gary N.S., Gelbart W.M., Glasser K.,
RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Itoigawa C.,
RA Jatali M., Kalush F., Karpen G.H., Ke Z., Kenton J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Moulton G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclib J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheefer F., Shen H.,
RA Shue B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T.,
RA Spter E., Spreading A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhu S., Zhu X., Smith H.O.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,

SQ SEQUENCE 407 AA; 46276 MW; 1A44765E3092BD07 CRC64;

Query Match 19.4%; Score 454; DB 2; Length 407;
Best Local Similarity 30.7%; Pred. No. 1.6e-24;
Matches 129; Conservative 84; Mismatches 151; Indels 56; Gaps 17;

QY 39 QKLSFTKLPLPLEDEIDEDTFHKSNNYSRAKAAPSIFGDVYNLAKQLVFITYDLFPKTMH 98
||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 35 QKICE---KKPL--LPCKDYEEAGNVYAIRKQSLISIQILD--GIIFAG-VFEGGLTH 84
||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 99 M-AVSLINAVLPFRHMVSIVAOSLCFLGSLSSISTLVDPPLSYSHFVEEKGFNKL 156
|| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 85 LEDLTHYN-LP---ETLGIVFALLFLAIQSVA---LPISITTYMHLDKEGEFSKV 135
|| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 157 TVOLMTDMIKSLTLAVAGGPILYLFLFKIPDKFPTDFLWYMF--LFVOOILAMTIP 214
::: | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 136 SLSLFFKNQFPGGLSTLTSGLLIYTLMILEHENV---WEISSFFVVFEMILANLEYP 192
::: | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

QY 215 VFIMPMENKPLPLEDEGLKKSIESLADEVPRPLDIPIVDISKSSNSMAYFTGLPFPSK 274
| : ||| : ||| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 193 K-IQLFPEEPPLPNRRDLESIEGMMDKVGKSEGISFIYMAASKRGRNALAFVFGIGL-KNK 250
| : ||| : ||| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :

QY 275 RIVLEDPLVNSNDDELTAVALHEIGHQMKHHVMVYFESQLHTPLSISLFTS----IYR 330
| : ||| : ||| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :

Db 251 RVVLFDPLTSVGVGEGLLALIGHGHKRNKDLSLGSIMGGLALVALPALIAHLPLPVE 310
| : ||| : ||| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :

QY 331 NTSFYNTGFELEKSTGSDVPVIYTKPEPPIIGFM-LFNDDLTLPLECAMQFVMSLIRTH 389
|| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 311 -----GNVSOPTASLI-----AILLLPLVESFTAMPL-----IGFSRKN 347
|| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 390 EYQADAVAKKLGKYKONICRALIDLQKNLSMTWNVDPLYSSHYHSPTLAERSTALDYSE 449
|| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 348 EYNADKFCGASLSRSKEVLAKALVSIYESNKAPPYSPHFVFLFHPTPLRLKALDYETE 407
|| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 9 PRELIMINARY; PRT; 447 AA.

Q9V7W6 AC Q9V7W6
Q9V7W6 ID Q9V7W6
AC Q9V7W6
DT 01-MAY-2000 (TREMBLrel.13, created)
DT 01-MAY-2000 (TREMBLrel.13, last sequence update)
DT 01-JUN-2001 (TREMBLrel.17, last annotation update)
DE CG9001 PROTEIN.
GN CG9001.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Empidoidea; Drosophilidae; Drosophila.
RX NCBI_TaxID=7227;
RN XN [1]
RP SEQUENCE FROM N.A.
RC SMPAIN-BERKELEY.
RE MEDLINE-20196006; PubMed-10731137;
RA Adams M.D., Celisner S.E., Holt R.A., Evans C.A., Gocayne J.D.,
Rana Anantides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
R Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
R Brandon R.C., Rogers Y.-H.C., Blatzej R.G., Champs M., Pfeiffer B.D.,
R Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
R Abill J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,
R Ballen R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
R Beeson K.Y., Benos P.V., Bernan B.P., Bhandari D., Bolshakov S.,
R Burkova K., Botchan M.R., Bouck J., Brockstein P., Brottier P.,
R Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
R Cherry J.M., Cavley S., Dahlke C., Davenport L.B., Davies P.,
R De Paulos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
R Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
R Durbin K.J., Evangelista C.C., Ferriz C., Ferreira S., Fleischmann W.,
R Foster C., Gabriellian A.E., Gary N.S., Gelbart W.M., Glasser K.,
R Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
R Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
R Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,

| | | | | |
|---------------------------|--------|--------------------|------------|-------------|
| Query Match | 17.4%; | Score 406.5; | DB 5; | Length 447; |
| Best Local Similarity | 28.0%; | Pred. NO. 4.1e-21; | | |
| Matches 132; Conservative | 78; | Mismatches 175; | Indels 87; | Gaps 17; |

| Q9HSA2 | PRELIMINARY; | PRT; | 428 AA |
|-----------|--------------|------|--------|
| ID Q9HSA2 | | | |

OS Q9H5K2; AC
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE ZINC METALLOCOPROTEINASE HOMOLOG.
GN CAAX OR VNG0329G.
OS Halodactylum sp. (strain NRC-1).
GN Halodactylum sp.

| | | | | |
|-----------------------|------------------|------------------|------------|-------------|
| Query Match | 16.1%; | Score 378; | DB 1; | Length 428; |
| Best Local Similarity | 29.3%; | Pred. NO. 4e-19; | | |
| Matches 110; | Conservative 64; | Mismatches 143; | Indels 58; | Gaps 12; |

| | |
|--------------|--------|
| Q9V7W5 | |
| ID | Q9V7W5 |
| PRELIMINARY; | PRT; |
| | 426 AA |

AC Q9VW05; 2000 (TrEMBLrel. 13, Created)
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
 DE CG9002 PROTEIN.
 GN CG9002.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachyera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxId=7227;
 RN [1]
 RZ SEQUENCE FROM N.A.
 RP STRAIN=BERKELEY;

RX MEDLINE-20196006; PubMed-10731132;
 RA Adams M.D., Celisner S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champs M., Pfeiffer B.D.,
 RA Man K.H., Doyle J., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfankoch C., Baldwin D.,
 RA Baller R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brothier P.,
 RA Butris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA De Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Dudin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwan C.,
 RA Jaisli M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spleter E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yen R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zhang X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,
 RT "The genome sequence of *Drosophila melanogaster*."
 RL Science 287:2185-2195(2000).
 DR EMBL: AE003804; AAF57924.1;
 DR FlyBase: FBgn0034174; CG9002.
 DR InterPro: IPR001915; Peptidase_M48.
 DR Pfam: PF01435; Peptidase_M48; 1.
 SQ SEQUENCE 426 AA; 49971 MW; 0760ECF84F291927 CRC64;

Query Match 15.9%; Score 372.5; DB 5; Length 426;
 Best Local Similarity 25.1%; Pred. No. 9.8e-19;
 Matches 110; Conservative 81; Mismatches 185; Indels 63; Gaps 13;

QY 33 LTVRGYOK-LSEFTKLPVLEDEIDDETHKSNYSRAKAKFSIPGDVYNLAKVFIKXD 91
 DB 3 LTRKQALVCLNAINWPEERGLIPEIYHRAHYELHTELDIMWYLLDLITLLELLIG 62
 QY 92 LEPKIMHNAVSLNANVLNVRFMHSTVAQSLGLSLSTLVLDPLSYSHVLEEKF 151
 DB 63 FYPLMWSLSATLQKITSGEIMI-----TLIFVYIYICIRPLVILYKCLELRY 116
 QY 152 GFN-KLTVQMLTMDIKSITLTAIGAIGPL-----YLPKIDKFPPTDPLWYIM 199
 DB 117 GMSGFPPYLYCCIGAMISLLSQLVLPALAAIVSVKFIGYF-----FLWF-W 165
 QY 200 VFLFVOLLAMTIPVFLIMPENKFTPLEDG-ELKKSSTESLADRVGFPLDKIFVLDGSR 258
 DB 166 LFWATFTLLVFLFPLPYCCPCIGROVVLPEGTALYMEKRVCDVVGFPKRVFIIR-TYT 224
 QY 259 SSSHNAVFTGLPFTSKRIYLFEDTLVNSNTD-----ETVLAHEIGHM 302
 DB 225 MQYSNAVYRG-SCCLKRIYIFDTLLNKGKENEIHPYEVGCLINIQVAGVCHLGM 283
 QY 303 OKNHTVNVNYSQHTPLFLFSITSYRNTSYFNFGFLEKSTGSEVDVITKEPPII 362
 DB 284 KHGHFYKATITMKIHFFITMGLGFFHSPOLYMAVGF-----EPGV---MPLIV 330

QY 363 GFML-FNDLTPLECAMQFVMSLSRTHGYADAVAKKIGYKONICRALID:QINKNLSM 421
 DB 331 GTITLKKALRPYTLANLVMLNLRREYADKFAHRNGYSIOLMALVKIYADHMSPP 390
 QY 422 NNDPLYSSYHVSHTPLAER 440
 DB 391 VYDQCYARHNNHTPIILQR 409
 RESULT 12
 ID 09VTH6 PRELIMINARY; PRT: 504 AA.
 AC 09VTH6:
 DT 01-MAY-2000 (TREMBLrel. 13, Created)
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE CG7573 PROTEIN.
 GN CG7573.
 OS *Drosophila melanogaster* (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Insecta; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; *Drosophila*.
 OX NCBI_TaxID=7227;
 [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BERKELEY;
 RX MEDLINE-20196006; PubMed-10731132;
 RA Adams M.D., Celisner S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champs M., Pfeiffer B.D.,
 RA Wan K.H., Doyle J., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfankoch C., Baldwin D.,
 RA Baller R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brothier P.,
 RA Butris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA De Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Dudin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwan C.,
 RA Jaisli M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spleter E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yen R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zhang X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,
 RT "The genome sequence of *Drosophila melanogaster*."
 RL Science 287:2185-2195(2000).
 DR EMBL: AE003545; AAF50072.1;
 DR FlyBase: FBgn0036153; CG7573.
 DR InterPro: IPR001915; Peptidase_M48.
 DR Pfam: PF01435; Peptidase_M48; 1.
 SQ SEQUENCE 504 AA; 57737 MW; 892022B675E6D49E CRC64;

| | | | | |
|-----------------------|------------------|--------------------|------------|-------------|
| Query Match | 10.3%; | Score 241.5; | DB 5; | Length 504; |
| Best Local Similarity | 20.7%; | Pred. NO. 2.3e-09; | | |
| Matches 93; | Conservative 85; | Mismatches 190; | Indels 81; | Gaps 17; |

| | | | |
|-----------|---|---|--------------|
| Qy | 28 | SFESELYTRYOKVOKISETLPLP-VLEDEIDDETFHKSNYSNAKAKFSIPGGVYNLAQKLV | 86 |
| Db | 85 | SFHLILCCROALKCKRISVPPKOMEGLISQEPASOKDOLHASSLEF-----NIAIDTL | 1400 |
| Qy | 87 | FIKYDLF---PKIMHNAVSLMLNAVLPVRFHMVSTVAOSLCFGLLSLSTLYDPLSY | 142 |
| Db | 141 | YSCDLVLYCTLAFLMKLVGM-----YHADSTWLNVTFMVVFSTYLVVRKLPSEFY | 192 |
| Qy | 143 | SHFYVEEFGFN-KLYVOLMTDMIKSLTILVAIGPILYFLKIFPKFPPLDL--WYIM | 199 |
| Db | 193 | EKLVLDPKRYNDPEKTPPL-LGICLALVFVAV-LEPLTALFMAT-----HMLSEMYFT | 243 |
| Qy | 200 | VELFEV---VOILAMTIIPVFIEMFNKFTPLEDEGELKSSIESIADVGFPLDKIFVIGD | 255 |
| Db | 244 | LEVMGLGLGSLVLIYALFGLVGLGVCCLCKSRKMNSTDMDNSLAKVALDDPNFP-GRVYAV-- | 300 |
| Qy | 256 | SKRSSHNAVYTGPLEFT-----SKRIYFDTL-----VNSNSTD-----EIT | 292 |
| Db | 301 | -----HTFVHGPRPTAWVMGCCCLRLDIDNLTKNRNGSSDDPFGMGAGLNDQELA | 353 |
| Qy | 293 | AVLAHEIGHMOKNHVNWVIFSQLHTFLIFSLFTSIYRNTSFYTRPQFPLEKSTGSPVDP | 352 |
| Db | 354 | AFVAHQHLHMOGLMHVAKGALITYLLLYLLLFETICNRMTTLVAAGF----- | 401 |
| Qy | 353 | VITKEEPIIIGFMELNDLTPPL-ECAMQFVNSLISRTHEYOADVAKKLGYNKONLCRALI | 411 |
| Db | 402 | --TFEYPSVGFWLVYKKLMPDIYHDISWIVFECIRHFEEAADAAYVNRGGLPMRALL | 459 |
| Qy | 412 | DLQIKNLSTWVDPLYSYHSHPTLAER | 440 |
| Db | 460 | KLFSDDYEEFYVDQCYLMMHRLRPSVLQR | 488 |
| RESULT 13 | | | |
| Q9V0G6 | | PRELIMINARY: | PRT: 338 AA. |
| AC | Q9V0G6; | | |
| DT | 01-MAY-2000 (TREMBlrel. 13, Created) | | |
| DJ | 01-MAY-2000 (TREMBlrel. 13, Last sequence update) | | |
| D7 | 01-JUN-2001 (TREMBlrel. 17, last annotation update) | | |
| DE | CAAX PRENTL PROTEASE 1 RELATED. | | |
| GN | PAB0555. | | |
| OS | Pyrococcus abyssi. | | |
| CC | Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus. | | |
| OX | NCBI_Taxid=29292; | | |
| RN | [1] | | |
| RN | SEQUENCE FROM N.A. | | |
| RP | STRAIN-ORSAY. | | |
| RC | | | |
| RA | Hellig R.; | | |
| RT | "Pyrococcus abyssi genome sequence: insights into archaeal chromosome structure and evolution". | | |
| RL | Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases. | | |
| DR | EMBL: AJ348285; CAB49737.1; -. | | |
| DR | Interpro: IPR001915; Peptidase_M48. | | |
| DR | Pfam: PF01435; Peptidase_M48; 1. | | |
| DR | Complete proteome. | | |
| QO | SEQUENCE 338 AA; 38790 MW; 35451FEB55F2AEB4 CRC64; | | |

```

Db      61  FDFAASFEPLODKIIVLPVLPVLSSTLITRELGKIRRGIVLAVF----- 110
QY      198  IMVLELFVY-----QLAMTIIPVFTMPENKF-----TPLEGECLKKSTESL 239
Db      111  -IIVLELFVLSFTIIPMYVIVYLLMFSLAEFLFSRIYRRFPHGTWM--SGELIRAKVEEL 168
QY      240  ADRYGFLPDKLFVINDGSKRSSHSAVETGLPFTSKRIVLDPDTLVNSNSTDEITAVLAHET 299
Db      169  CRRANVANVEEVIYID---EERIGAPVTGM--KKGITFTIKGAIEKLNEMELLAVIAHEL 222
QY      300  GHMKKNHIVNVAIVISQLHTFLFELFISYIRNITSFYNTPGFPLEKSTGSDVDPYITKEFP 359
Db      223  GHVYRRHALKREL-----ALVFEL-----SLP 244
QY      360  IIGFMLEFNDLPLELCAMQFVMSL-----ISRTHEYQADAYAKKLGKYKONLCRAL 410
Db      245  -IIGYMGQDI--PL--FLSFVMSIALVPESTQIAKKFEIADRFAASLVGPINVIKAL 299
QY      411  IDLQIKNLSNTMNVDPKYSYHYSHPTLAESTDAID 445
Db      300  EKVYTKEGLPKRTPRWYNIIH--SPHSLEERRALE 333

```


Db 174 GNFYDKALKKNEPDCIGYFATIFAEV--LGLIASTIVMFSRRREFRADAGAHLAG 231
 QY 403 KONLCRALIDQ-----IKNLSTMNVDPLYSSTHY-----SPTLAERSTAL 444
 Db 232 TGAMIAALQRLRSOGVGVPMPTLMAFGINGLKHGLAGLMSHPLEDRIEAL 286

RESULT 15

O9JYV19
 ID O9JYV19 PRELIMINARY: PRT: 279 AA.
 AC O9JYV19;
 DT 01-OCT-2000 (Tremblrel, 15, Created)
 DT 01-OCT-2000 (Tremblrel, 15, Last sequence update)
 DT 01-JUN-2001 (Tremblrel, 17, Last annotation update)
 DE PUTATIVE MEMBRANE-BOUND ZINC METALLOPEPTIDASE (EC 3.4.24.).
 GN HTPX OR NMA1031.
 OS Neisseria meningitidis (serogroup A).
 OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
 OX NCBI_Taxid=65699;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=22491 / SEROGROUP A / SEROTYPE 4A;
 RX MEDLINE=20222556; PubMed=10761919;
 RA Parkhill J., Achtman M., James K.D., Bentley S.D., Churcher C.,
 Klee S.R., Morelli G., Basham D., Brown D., Chillingworth T.,
 Davies R.M., Davis P., Devlin K., Feltwell T., Hamlin N., Holroyd S.,
 Jorgels K., Leather S., Moule S., Mungall K., Quail M.A.,
 Rajandream M.A., Rutherford K.M., Simmonds M., Skelton J.,
 Whitehead S., Spratt B.G., Barrall B.G.;
 RT "Complete DNA sequence of a serogroup A strain of Neisseria
 meningitidis 22491".
 RT Nature 404:502-506(2000).
 RL EMBL: AL162754; CAB84300.1;
 DR EMBL: AL162754; CAB84300.1;
 DR InterPro: IPR001915; Peptidase_M48.
 DR Pfam: PF01435; Peptidase_M48; 1.
 DR PROSITE: PS00142; ZINC_PROTEASE; UNKNOWN_1.
 KW Hydrolyase; Complete proteome.
 SQ SEQUENCE 279 AA; 30174 MW; 713442CD6CAC015E CRC64;

Query Match 6.3%; Score 148.5; DB 2; Length 279;
 Best Local Similarity 25.7%; Pred. No. 0.0044;
 Matches 59; Conservative 42; Mismatches 88; Indels 41; Gaps 10;

QY 228 EDGELKKSIESLADRVGFPDLKIFVIDGSKRSSHNAFTGIPFTSKRIVLFDLTVNSNS 287
 Db 75 EEAMLMTVEAARQWNLKTPVAIY---HSPEPNAFATGASRNSSLIAVSTGLDHT 130
 QY 288 TDEITAVIAHEIGHMOKHIVNM-VIFSQLTFLIFSLSIYRNTSYNTFGFFLEKST 346
 Db 131 RDEVEAVLAHEMAHYGNDMTLLTLOGVNTFYVF--LSRIIANLIRNNDG---SOSQ 185
 QY 347 GSFVDPVTKKEPFIIGFMLENDLTLPLECAMQFVMSLSIRTHGQADAVAKKLGKXNL 406
 Db 186 GTYF--LVSMYFQILFGFL-----ASLIVMFSRQREYRADAGAKLVGAPKM 231
 QY 407 CRAL-----IDQIKNLSTMNV----DPLVSSYHSHPTLAERSTAL 444
 Db 232 ISALQRLKGNFVDLP--EEMNMGIAGTDRDLSL---THPSLDNRIRAL 276

Search completed: February 12, 2002, 09:32:01
 Job time: 348 sec

THIS PAGE BLANK (1.3PTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 12, 2002, 09:25:34 ; Search time 21.66 Seconds
(without alignments)
470.637 Million cell updates/sec

Title: US-09-165-460a-2
Perfect score: 2342
Sequence: 1 MEDLKTILDPNIPMKLIIS.....HPTLAEKSTALDYVSEKKK 453

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA.*
1: /cgn2-6/ptodata/2/1aa/5A-COMB.pep.*
2: /cgn2-6/ptodata/2/1aa/5B-COMB.pep.*
3: /cgn2-6/ptodata/2/1aa/6A-COMB.pep.*
4: /cgn2-6/ptodata/2/1aa/6B-COMB.pep.*
5: /cgn2-6/ptodata/2/1aa/PTUS-COMB.pep.*
6: /cgn2-6/ptodata/2/1aa/Backfile1.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|-------------------|
| 1 | 653 | 27.9 | 474 | 3 | US-09-022-699-2 |
| 2 | 105 | 4.5 | 3200 | 2 | US-08-477-451-8 |
| 3 | 100.5 | 4.3 | 1226 | 2 | US-08-540-804-12 |
| 4 | 100.5 | 4.3 | 1226 | 2 | US-08-218-265-12 |
| 5 | 100.5 | 4.3 | 1226 | 3 | US-08-521-872-12 |
| 6 | 100.5 | 4.3 | 1226 | 4 | US-08-590-399-12 |
| 7 | 94 | 4.0 | 1876 | 2 | US-08-619-554-2 |
| 8 | 93.5 | 4.0 | 485 | 2 | US-08-477-451-22 |
| 9 | 92.5 | 3.9 | 284 | 1 | US-08-118-270-67 |
| 10 | 92.5 | 3.9 | 284 | 5 | PCT-US93-08528-67 |
| 11 | 92.5 | 3.9 | 1030 | 4 | US-09-091-117-2 |
| 12 | 92 | 3.9 | 1285 | 1 | US-07-582-945-2 |
| 13 | 92 | 3.9 | 1285 | 2 | US-08-453-141-2 |
| 14 | 92 | 3.9 | 1285 | 3 | US-08-293-314-2 |
| 15 | 90 | 3.8 | 485 | 1 | US-07-991-8678-42 |
| 16 | 90 | 3.8 | 485 | 2 | US-08-544-332-42 |
| 17 | 89 | 3.8 | 194 | 4 | US-08-959-004-9 |
| 18 | 88.5 | 3.8 | 498 | 1 | US-08-457-274A-24 |
| 19 | 88.5 | 3.8 | 498 | 5 | PCT-US95-05758-24 |
| 20 | 88.5 | 3.8 | 1786 | 2 | US-08-477-451-16 |
| 21 | 88 | 3.8 | 1045 | 1 | US-07-596-467-6 |
| 22 | 88 | 3.8 | 1045 | 1 | US-07-934-374-6 |
| 23 | 88 | 3.8 | 1045 | 1 | US-07-783-861C-6 |
| 24 | 87.5 | 3.7 | 1781 | 2 | US-08-477-451-11 |
| 25 | 87 | 3.7 | 440 | 1 | US-08-307-499-15 |
| 26 | 87 | 3.7 | 440 | 4 | US-09-299-268-15 |
| 27 | 87 | 3.7 | 599 | 1 | US-08-295-814E-11 |

| | | | | | | |
|----|------|-----|------|---|-------------------|--------------------|
| 28 | 87 | 3.7 | 599 | 1 | US-08-240-783B-4 | Sequence 4, Appl1 |
| 29 | 87 | 3.7 | 599 | 3 | US-09-084-813-4 | Sequence 4, Appl1 |
| 30 | 87 | 3.7 | 599 | 4 | US-09-343-361-11 | Sequence 11, Appl1 |
| 31 | 87 | 3.7 | 599 | 5 | PCT-US92-09662-4 | Sequence 4, Appl1 |
| 32 | 86.5 | 3.7 | 502 | 1 | US-08-484-840-3 | Sequence 3, Appl1 |
| 33 | 86.5 | 3.7 | 502 | 1 | US-08-483-094-3 | Sequence 3, Appl1 |
| 34 | 85 | 3.6 | 599 | 1 | US-08-301-722A-5 | Sequence 5, Appl1 |
| 35 | 84 | 3.6 | 432 | 4 | US-09-370-253-12 | Sequence 12, Appl1 |
| 36 | 84 | 3.6 | 1394 | 4 | US-09-213-053-2 | Sequence 2, Appl1 |
| 37 | 83.5 | 3.6 | 277 | 1 | US-08-118-270-68 | Sequence 68, Appl1 |
| 38 | 83.5 | 3.6 | 277 | 5 | PCT-US93-08528-68 | Sequence 2, Appl1 |
| 39 | 83.5 | 3.6 | 1956 | 4 | US-08-843-417-2 | Sequence 2, Appl1 |
| 40 | 83 | 3.5 | 1079 | 3 | US-09-136-652-2 | Sequence 4, Appl1 |
| 41 | 83 | 3.5 | 1895 | 2 | US-08-619-554-4 | Sequence 2, Appl1 |
| 42 | 82.5 | 3.5 | 355 | 2 | US-08-458-970A-2 | Sequence 2, Appl1 |
| 43 | 82.5 | 3.5 | 1079 | 1 | US-08-455-343A-55 | Sequence 55, Appl1 |
| 44 | 82.5 | 3.5 | 1079 | 2 | US-08-223-305C-55 | Sequence 2, Appl1 |
| 45 | 82 | 3.5 | 620 | 1 | US-08-301-722A-2 | Sequence 2, Appl1 |

ALIGNMENTS

RESULT 1
US-09-022-699-2
Sequence 2, Application US/09022699
Patent No. 6060277
GENERAL INFORMATION:
APPLICANT: KIKLY, KRISTINE
APPLICANT: SOUTHAN, CHRISTOPHER
APPLICANT: KNAB, ANNE
TITLE OF INVENTION: Human AFCl
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESSES:
ADDRESSEE: RATNER & PRESTIA
STREET: P.O. BOX 980
CITY: VALLEY FORGE
STATE: PA
COUNTRY: USA
ZIP: 19482
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/022.699
FILING DATE: 12-FEB-1998
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 97304440.7
FILING DATE: 12-JUN-97
ATTORNEY/AGENT INFORMATION:
NAME: PRESTIA, PAUL F
REGISTRATION NUMBER: 23,031
REFERENCE/DOCKET NUMBER: GH-70380
TELECOMMUNICATION INFORMATION:
TELEPHONE: 610-407-0700
TELEFAX: 610-407-0701
TELEX: 846169
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 474 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-022-699-2
Query Match 27.9%; Score 653; DB 3; Length 474;
Best Local Similarity 34.7%; Pred. No. 1.1e-59;
Matches 161; Conservative 73; Mismatches 160; Indels 62; Gaps 8;

```

QY 22 FSIQSFESFSLYRQ---YQKLSFTKLPVLEDEIDETPHKSRNYSRAKAKSIFCDV 78
  || : : : : : || : : : : : || : : : : : || : : : : : || : : : : :
Db 26 FSWTYLMEFTFLAQRORRYK--TTTHVPPELGQIMDSFTFKSRUYQDKTFTFSWSGL 83
QY 79 YNLQKLVYIKIDLPFKTKHMAVSLNNAVLPVRPHMVSIVASQSLCFLGLSSLSLTVLDP 138
  || : : : : : || : : : : : || : : : : : || : : : : : || : : : : :
Db 84 YSETGCTLLFLFGIPYLMRLSGRCGY---AGFGPEYITQSLVFLLLATLFSALTGLP 140
QY 139 LSYSHFVLEEKFGFKLTVOLMTDMIKSLTLAVAGSPILYLFKTIIDKRPPTPLMWI 198
  || : : : : : || : : : : : || : : : : : || : : : : : || : : : : :
Db 141 WSLYNTFVEEKHGNQQLTGFMDAKKVVYTOCILLPVSSLLYTIKIGDYFTYA 200
QY 199 WVEFLVVOIAMTIIIPVIMPENKFTPLEDELKKSIESLADRYGFPDLKIFVIDGSR 258
  || : : : : : || : : : : : || : : : : : || : : : : : || : : : : :
Db 201 WLFYLVSLVLTIVADYIAPLEFKFTPLPECKLKEIEVMKSIDPFLTKYVVEGSR 260
QY 255 SSHSNAYFTGLPFTSKRYLFTLVNSN----- 287
  || : : : : : || : : : : : || : : : : : || : : : : : || : : : : :
Db 261 SSHSNAYFTGF-FKNKRIVLPDTLLEYSVLNKDIOEDSGMEPRNEEGNSEIRAKYKN 319
QY 288 -----TDEITAVLAHEIGHMOKNHVNVIFSOHLTFLIFSLFTSIYNTSFYNTFGF 341
  || : : : : : || : : : : : || : : : : : || : : : : : || : : : : :
Db 320 KQGGCKNEVLAVLGHGSHW-KWDIOSKISLAREPFVFLVAVLGKRLFAAFGRY 378
QY 342 LEKSTGSPVDYPTKEFFPIIGFM-LFNDLTPLECAMQFVMSLSIRTHEYOADAVAKKL 400
  || : : : : : || : : : : : || : : : : : || : : : : : || : : : : :
Db 379 DSQ-----PFLIGLLIIFQIFSPYNEVLSFCITVLSRRREFQADAFAKKL 424
QY 401 GYKQMLCRALDLOIKNSTMNVDPLYSYHSHPTLAERSTAL 444
  || : : : : : || : : : : : || : : : : : || : : : : : || : : : : :
Db 425 GKAKDLYSALIKLNKDNLGFVSDWLFSMWHYSHPPLLERLOAL 468

```

RESULT 2

```

US-08-477-451-8
: Sequence 8, Application US/08477451
: Patent No. 5928865
: GENERAL INFORMATION:
: APPLICANT: Covacci, Antonello
: TITLE OF INVENTION: Helicobacter Pylori CagI Region
: NUMBER OF SEQUENCES: 46
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Chiron Corporation
: STREET: 4560 Horton Street
: CITY: Emeryville
: STATE: CA
: COUNTRY: USA
: ZIP: 94608-2916
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/477,451
: FILING DATE: 07-JUN-1995
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: McClung, Barbara G.
: REGISTRATION NUMBER: 33,113
: REFERENCE/DOCKET NUMBER: 0335, 002
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 510-601-2708
: TELEFAX: 510-655-3542
: INFORMATION FOR SEQ ID NO: 8:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 3200 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: US-08-477-451-8

```

```

Query Match 4.5%; Score 105; DB 2; Length 3200;
Best Local Similarity .24, 8%; Pred. No. 0.25;
Matches 102; Conservative 55; Mismatches 152; Indels 102; Gaps 23;

QY 38 YQKLSFTKLPVLEDEIDETPHKSRNY---SRAKAKFSI-----FGDY-----N 80
  || : : : : : || : : : : : || : : : : : || : : : : : || : : : : :
Db 1274 FRNKSNTIIPGVISLRVLEIFNLSEFFLFLAQVYSIILCKFHEIYFVLDLFLICS 1333
QY 81 LAQKLVF---IK-----YDLFPKIHMAVS-----LNAVLPV----- 110
  || : : : : : || : : : : : || : : : : : || : : : : : || : : : : :
Db 1334 LACELVFEVEIKTRVNPFLTGYDA-TAVFHMHNCSHIVALLVFTINGVPIPSGR 1392
QY 111 -----RHHVSTY---AQSLCFLGLSSLSLTVLDPVLSYSHFVLEEKGFNKL 156
  || : : : : : || : : : : : || : : : : : || : : : : : || : : : : :
Db 1393 TTIAPIQNEFFRYALIMPVHRRQSV-IYNQILELVIGINVSDSLNHHT- AKSEFRKI 1450
QY 157 TVQWITDMIKSLTLAVAGSPILYLFKIPKPTDFIMYIMVFLVVOIAMTII-IP 214
  || : : : : : || : : : : : || : : : : : || : : : : : || : : : : :
Db 1451 TPEI-----KISFRALYVGGSGGYIMDMGMWFLIDCFDIR-RLFFVKYKKNIQPIP 1505
QY 215 VIMPENKFTPLEDELKKSIESL-ADRYG-----PLDKIFVID--GSKR 258
  || : : : : : || : : : : : || : : : : : || : : : : : || : : : : :
Db 1506 TSRLSIFFCFLPFFOGLIKHAHELLIRSDLIPIYNDCKRFLGFRQKLRFLTIKRGLYD 1565
QY 259 SSHSNAYFTGLPFTSKRYLFTLVNSNSTDEITAVLAHEIGHMOKNHVNVIFSOHLT 318
  || : : : : : || : : : : : || : : : : : || : : : : : || : : : : :
Db 1566 IFHKISDSGVDF--RIVLMGHPCENSVTLIPKFLALCGRYINEIWEVFDIALLIDK 1623
QY 319 FLIFSLFTSIYNTSFYNTFGFPLEKSTGSPVDYPTKEFFPIIGFMFLND 369
  || : : : : : || : : : : : || : : : : : || : : : : : || : : : : :
Db 1624 VKILLFGRIF--DRFIGVF--PPLRDSFKSL--PIV-----LIND 1661

```

RESULT 3

```

US-08-540-804-12
: Sequence 12, Application US/08540804
: Patent No. 5919666
: GENERAL INFORMATION:
: APPLICANT: Young, Richard A.
: APPLICANT: Koleske, Anthony J.
: APPLICANT: Thompson, Craig M.
: APPLICANT: Chao, David M.
: TITLE OF INVENTION: No. 5919666el Factors Which Modify Gene
: TITLE OF INVENTION: Transcription and Methods of Use Therefor
: NUMBER OF SEQUENCES: 39
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
: STREET: Two Millitia Drive
: CITY: Lexington
: STATE: Massachusetts
: COUNTRY: USA
: ZIP: 02173
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/540,804
: FILING DATE: 11-OCT-1995
: CLASSIFICATION: 424
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/521,872
: FILING DATE: 21-AUG-1995
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/218,265
: FILING DATE: 25-MAR-1994
: ATTORNEY/AGENT INFORMATION:
: NAME: Granahan, Patricia
: REGISTRATION NUMBER: 32,227
: REFERENCE/DOCKET NUMBER: WHI94-03A2

```

TELECOMMUNICATION INFORMATION:

TELEPHONE: 617-861-6240

TELEFAX: 617-861-9540

INFORMATION FOR SEQ ID NO: 12:

SEQUENCE CHARACTERISTICS:

LENGTH: 1226 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-540-804-12

Query Match 4.3%; Score 100.5; DB 2; Length 1226;
Best Local Similarity 19.5%; Pred. No. 0.18;

Matches 96; Conservative 86; Mismatches 206; Indels 105; Gaps 21;

```
QY 7 ILDHNINPKLLISGFSAQSFESYLRYROYKLSYKLPVLEDEIDDEFHKSRYNS 66
DB 374 IIRKPTIRKLLISGLYLQDSNDRFV--HVQLINLKISPLMSQYN---MYLRNVM 426
QY 67 RAKAFSIFGDVYNLAOKLVFIKYDLFPKIMHAYS---LNAVLPVRHMVSTVAQSLC 123
DB 427 EVDVAFY--EIFNFDQ-LVEITBOIKRILSNDITNQLSTPLSIKIMVAEWLSHLC 482
QY 124 FLGLISLSTVLDL-----PLSYSHFVLEEKFGFNKLTVOAMITDMIKSLTAVAI 175
DB 483 -SGILSSVNRVYLLKIFKICIDLEVFNHF-----FKWIEFIVYHQLSDIESLEALM-- 534
QY 176 GGPILYLKIFDKFPPTFLMYIMFLVYV-QILAMTIIPV---IMPEN-----KF 224
DB 535 -DILLCYOKLESQFINHILFTKTFIFLYKKVLEKEDVPAVNTVSFPMFEMKFNKPNPF 592
QY 225 TPLEDGELKKSIESLADRVGFPDLKI-----FVIDSKRSSHNAFTGLPPTS 273
DB 533 VLKVNDRIRIELQSYNDEKLTETKLNKDSVULKVYSMINNSNOAVGQTNPFPEVOYN 652
QY 274 KRIVLEDT-LVNSNST-----DEITAVLAHEIGHQWKNHIVNVIQSOLHTF----- 319
DB 653 IREFLHNSIIDTNTSKQOKARRNVMLLIATNL----KEYIKFMSIFLKRKDFTNKNI 708
QY 320 -----LIFSLETSIYRNTSFYNTFGFLEKSTGSGF-----VDPVITKEPPIIGFM 365
DB 709 QLISKLTLFEVTONVLEVIIRLLPINLENNDGSGYGLFLKYHKEQFIKSNFEKI---- 764
QY 366 LFNDDLTPL-----ECAMQFVMSLSIRTHEQADAVAKKLCYKONLCALDLOIK 416
DB 765 ----LITCYELEKKYHNGDECIYVEILKILITTYGSS--PKLLATSTKIIMLLNDSVE 818
QY 417 NLSTNMVDPLEYSS 429
DB 819 NSSNLEIDILYYS 831
```

RESULT 4

US-08-218-265-12

Sequence 12, Application US/08218265

Patent No. 5922585

GENERAL INFORMATION:

APPLICANT: Young, Richard A.

APPLICANT: Koleske, Anthony J.

APPLICANT: Thompson, Craig M.

TITLE OF INVENTION: No. 5922585el Factors Which Modify Gene

NUMBER OF SEQUENCES: 35

CORRESPONDENCE ADDRESSES:

ADDRESS: Hamilton, Brook, Smith & Reynolds, P.C.

STREET: Two Millia Drive

CITY: Lexington

STATE: MA

COUNTRY: US

ZIP: 02173

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/218, 265

FILING DATE: 25-MAR-1994

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Granahan, Patricia

REGISTRATION NUMBER: 32,227

REFERENCE/DOCKET NUMBER: WHI94-03

TELECOMMUNICATION INFORMATION:

TELEPHONE: 617-861-6240

TELEFAX: 617-861-9540

INFORMATION FOR SEQ ID NO: 12:

SEQUENCE CHARACTERISTICS:

LENGTH: 1226 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-218-265-12

Query Match 4.3%; Score 100.5; DB 2; Length 1226;
Best Local Similarity 19.5%; Pred. No. 0.18;

Matches 96; Conservative 86; Mismatches 206; Indels 105; Gaps 21;

```
QY 7 ILDHNINPKLLISGFSAQSFESYLRYROYKLSYKLPVLEDEIDDEFHKSRYNS 66
DB 374 IIRKPTIRKLLISGLYLQDSNDRFV--HVQLINLKISPLMSQYN---MYLRNVM 426
QY 67 RAKAFSIFGDVYNLAOKLVFIKYDLFPKIMHAYS---LNAVLPVRHMVSTVAQSLC 123
DB 427 EVDVAFY--EIFNFDQ-LVEITBOIKRILSNDITNQLSTPLSIKIMVAEWLSHLC 482
QY 124 FLGLISLSTVLDL-----PLSYSHFVLEEKFGFNKLTVOAMITDMIKSLTAVAI 175
DB 483 -SGILSSVNRVYLLKIFKICIDLEVFNHF-----FKWIEFIVYHQLSDIESLEALM-- 534
QY 176 GGPILYLKIFDKFPPTFLMYIMFLVYV-QILAMTIIPV---IMPEN-----KF 224
DB 535 -DILLCYOKLESQFINHILFTKTFIFLYKKVLEKEDVPAVNTVSFPMFEMKFNKPNPF 592
QY 225 TPLEDGELKKSIESLADRVGFPDLKI-----FVIDSKRSSHNAFTGLPPTS 273
DB 533 VLKVNDRIRIELQSYNDEKLTETKLNKDSVULKVYSMINNSNOAVGQTNPFPEVOYN 652
QY 274 KRIVLEDT-LVNSNST-----DEITAVLAHEIGHQWKNHIVNVIQSOLHTF----- 319
DB 653 IREFLHNSIIDTNTSKQOKARRNVMLLIATNL----KEYIKFMSIFLKRKDFTNKNI 708
QY 320 -----LIFSLETSIYRNTSFYNTFGFLEKSTGSGF-----VDPVITKEPPIIGFM 365
DB 709 QLISKLTLFEVTONVLEVIIRLLPINLENNDGSGYGLFLKYHKEQFIKSNFEKI---- 764
QY 366 LFNDDLTPL-----ECAMQFVMSLSIRTHEQADAVAKKLCYKONLCALDLOIK 416
DB 765 ----LITCYELEKKYHNGDECIYVEILKILITTYGSS--PKLLATSTKIIMLLNDSVE 818
QY 417 NLSTNMVDPLEYSS 429
DB 819 NSSNLEIDILYYS 831
```

RESULT 5

US-08-521-872-12

Sequence 12, Application US/08521872

Patent No. 6015682

GENERAL INFORMATION:

APPLICANT: Young, Richard A.

APPLICANT: Koleske, Anthony J.

APPLICANT: Thompson, Craig M.

APPLICANT: Chao, David M.


```

Db 483 -SGILSVNRVTLKIFKICIDLEVFHMF-----FKWIEFTVHQSLSDIESLEALM-- 534
Qy 176 GGPILYFLFKIPDKPEPTDFLWYIMVLEFVY-OILAMTIPVF-----IMPMEF-----KF 224
Db 535 --DILCYOKLTSOFINDHILFTKTFIFLYKKVLEKEDVPAYNVTSFEMFWMFKMKNPF 592
Qy 225 TPLEDEGLKKSIESLADRVGFPLDKI-----FVIDGSKRSSHSHNAFTGLPFTS 273
Db 593 VLKVNNDLRIELOSIVNDEKLEKTKNDKSEVLKYSMINSNQAVGTWNPPEVFGQN 652
Qy 274 KRIVEFDT-LVNSNST-----DETVLAHEIGHWOKNHIYVNSQLHTF----- 319
Db 653 IFFLLHNSIEDINTSKOFKARNVMMLIATNL-----KEYIKFMSIFLKRDFTNKMLI 708
Qy 320 -----LIFSLSIYRNTSFYNTGFELEKSTGSF-----VDPVITKKEPIIIGFM 365
Db 709 QILSLKLFEEVTONVGLYEITRIPLPINLENDGSGYGLFKYHKEQFKSFEKI----- 764
Qy 366 LFNDLITPL-----ECAMQFVMSLISRTHEYOADAAYAKKGYKONLCRALIDLOIK 416
Db 765 ----LLTCELEKKYHGNCEINYEIILKILITYGSS--PKLATSTKIIMLLINDSVE 818
Qy 417 NLTSMNVDPYSS 429
Db 819 NSSNILEDILYXS 831

```

RESULT 7

US-08-619-554-2

; Sequence 2, Application US/08619554

; Patent No. 5821353

; GENERAL INFORMATION:

; APPLICANT: DOUGLAS, Cameron M.

; APPLICANT: CHREBET, Gary L.

; APPLICANT: CLEMAS, Joseph

; APPLICANT: EL-SHERBEINI, Mohammed

; APPLICANT: FOOR, Forrest

; APPLICANT: KAHN, Jennifer

; APPLICANT: KELLY, Rosemarie, - PARENT, S.A.

; APPLICANT: MARRINAN, Jean, - RAMADAN, N.M.

; APPLICANT: MORIN, Nancy, - REGISTER, E.A

; APPLICANT: ONISHI, Janet, - SHEI, Gan-Ju

; TITLE OF INVENTION: DNA ENCODING 1,3 BETA-D GLUCAN

; TITLE OF INVENTION: SYNTASE SUBUNITS

; NUMBER OF SEQUENCES: 8

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: JOSEPH A. COPPOLA - MERCK & CO., INC.

; STREET: 126 EAST LINCOLN AVENUE - P.O. BOX 2000

; CITY: RAHWAY

; STATE: NJ

; COUNTRY: USA

; ZIP: 07065

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette

; COMPUTER: IBM Compatible

; OPERATING SYSTEM: DOS

; SOFTWARE: FastSeq for Windows Version 2.0

; CURRENT APPLICATION DATA: US/08/619.554

; FILING DATE: 01-AUG-1996

; CLASSIFICATION: 536

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER:

; FILING DATE:

; ATTORNEY/AGENT INFORMATION:

; NAME: COPPOLA, JOSEPH A

; REGISTRATION NUMBER: 38,413

; REFERENCE/DOCKET NUMBER: 19104PI

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 732-594-6734

; TELEFAX: 732-594-4720

; INFORMATION FOR SEQ ID NO: 2:

```

; SEQUENCE CHARACTERISTICS:
; LENGTH: 1876; amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
US-08-619-554-2

```

```

Query Match 4.0%; Score 94; DB 2; Length 1876;
Best Local Similarity 19.8%; Pred. No. 1.6;
Matches 79; Conservative 57; Mismatches 128; Indels 134; Gaps 18;

```

```

Qy 3 DKLTIIDHPNIPMKLII-SGFSIAQSFSESYLYTQYOKLSTKLPVLEDEIDDETFK 61
Db 392 DLNQLFWYPEGIAKIVLEDEGTRKILIEPLE-----ERYLRGD-----VMDVFKTYKE 441
Qy 62 SRNYSRAKAKFSIFGDVYNLAQKLVFIKYDLPKIMHNAVSIL-----NAV 107
Db 442 TRTWLHLVTNEN-----RIWMHISIFMWYFAVNSPTFYTHNQ 480
Qy 108 LPRFHWSTVAQSLCFIG-----LLSLSTLYVDLPLSYSHFVLEKFGFKLTVOLMT 162
Db 481 QLVNDQPLAAYKMACALGCTVASLIQIVATLCEN-----SEVPRKAGAOHLSRRWF 534
Qy 163 TDMIKSLTAAVAGGPILYFLKIPDKPEPTDFLWYIMVLFVYOLIAMTIIIPVFIIMPEN 222
Db 535 LCTIFGINL-----GPILFVFAVDKQTVYSTAHHVAAMVFVA--ATIIFFSTIMPUG 587
Qy 223 KFTPLEDELKKSIESLADRVGFPLDKIFVIDGSKRSSHSHNAFTGLPFTSKRIYLPFTL 282
Db 588 LFT-----SYMKS-----TRRYASQT- 605
Qy 283 VNSNSTDEITAVLA--HEIGHWOKNHIYVMVIF-----SOLHTFLIFS-----LFTSIY 329
Db 606 -----FTAAFAFDLHGIDRWM-SYLWVTVYFAKYSBSYFVLSIRDPIRLISTAM 656
Qy 330 RNTSFYNTGFELEKSTGSFVDPVITKERPIIIGMFL 367
Db 657 RCTGEY-WMGAVLCK-----VQPKIVGLVIAITDFILF 688

```

RESULT 8

US-08-477-451-22

; Sequence 22, Application US/08477451

; Patent No. 5928865

; GENERAL INFORMATION:

; APPLICANT: Covacci, Antonello

; TITLE OF INVENTION: Helicobacter pylori CagI Region

; NUMBER OF SEQUENCES: 46

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Chiron Corporation

; STREET: 4560 Horton Street

; CITY: Emeryville

; STATE: CA

; COUNTRY: USA

; ZIP: 94608-2916

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA: US/08/477.451

; FILING DATE: 07-JUN-1995

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: McCung, Barbara G.

; REGISTRATION NUMBER: 33,113

; REFERENCE/DOCKET NUMBER: 0335.002

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 510-601-2708

; TELEFAX: 510-655-3542

INFORMATION FOR SEQ ID NO: 22:
SEQUENCE CHARACTERISTICS:
LENGTH: 485 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-477-451-22

Query Match 4.0%; Score 93.5; DB 2; Length 485;
Best Local Similarity 19.8%; Pred. No. 0.24;
Matches 93; Conservative 63; Mismatches 159; Indels 155; Gaps 24;

QY 7 LIDHPNIPWKLI-----SGFSIAQSFESYLTROYOKLSETKLPVLEDEIDETPH 60
DB 49 LLSHAKFTMKLIILMTQOMNMFVQ---KFLKTRREISLFPNISIKPLKTKIFLSAVT 105
QY 61 KSRN-----YSRAKAKF--SIFG---DYVNL-----AOKLVEIKY---DLPPK--- 95
DB 106 KLAFLKPKPIYFWYSNPKPFIVSVFRCICVSNLFLIMWFSEKLIKICYNKIKRIM 165
QY 96 -IMHNAVSLINAVLPVRFMVSTVAA--SLCFLGLISLSTLVDPPLSYSHFVLEKFG 152
DB 166 LVMWONIFVCDWMLIAFHESYQIOEFLVCYCEMLPTKTRIRD-P----- 210
QY 153 FPKLTV-QLMTDMIKSLTAVAGSPILYFLKIFDKRPDPLWYIMVFLVQILAMT 211
DB 211 -NKOETOPKIKGLIMKTLASLLOG-----TNLEFTGISLDFSMILFLF----- 256
QY 212 IIPVIMPNKFTPLEDELKKSIESLADRVGFPLDKI-EVIDGSKRSHSNVFTGL 269
DB 257 -----FLMLMG-----LMDALGKKFNLPMDNIKMAEVLKNG----- 290
QY 270 PFTSKRIVFQRLVNSNSTDEITAVLAHIGMOKNHIYNAVIFSOHLFLFLSFTSY 329
DB 291 -----FDSIKNMG-----ALVNGFGSSKSDKTANKMSQVRLMCINYLKPKF 336
QY 330 R-----NTSFYNTFGF-----FLEKSTGSPVDPIYTK-----EFPII--TG 363
DB 337 HLSRALFQQRNRLLAIEYDITDSFLKM---SYLDPFLKIFITTDLAFLACLPFLKRLG 393
QY 364 FNLNDLTLPLECAMQF-----VMSLISRTHEYQADY 396
DB 394 FFLHRSTISIQVAIFRNPNIRGSTSPDALISVMSVSSONPPTFIY 443

RESULT 9
US-08-118-270-67
Sequence 67, Application US/08118270
Patent No. 5508384

GENERAL INFORMATION:
APPLICANT: Murphy, Randall B.
APPLICANT: Schuster, David I.
TITLE OF INVENTION: POLYPEPTIDES OF G-COUPLED PROTEIN
NUMBER OF SEQUENCES: 348
CORRESPONDENCE ADDRESS:
ADDRESSEE: BROWDY AND NEIMARK
STREET: 419 Seventh Street, N.W., Suite 300
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/118,270
FILING DATE: 09-SEP-1993
PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/943,236
FILING DATE: 10-SEP-1992
ATTORNEY/AGENT INFORMATION:
NAME: Townsend, Kevin G.
REGISTRATION NUMBER: 34,033
REFERENCE/DOCKET NUMBER: MURPHY-2A
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-628-5197
TELEFAX: 202-737-3528
TELEX: 248633
INFORMATION FOR SEQ ID NO: 67:
SEQUENCE CHARACTERISTICS:
LENGTH: 284 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-118-270-67

Query Match 3.9%; Score 92.5; DB 1; Length 284;
Best Local Similarity 21.1%; Pred. No. 0.14;
Matches 62; Conservative 45; Mismatches 96; Indels 91; Gaps 14;

QY 7 LIDHPNIPWKLIIGFSIAQSFESYLTROYOKLSETKLPVLEDEIDETPHKSRNY 65
DB 25 LLDHILHPTMYFLNLSADLCFSSKLQNMQ---SQVPSI-----PY 66
QY 66 SRAKAK---FSIFGVYNLAQKLVPIKYDLF-----PKIMHNAVSLINAVL 108
DB 67 ACGLAQIYFELFPFGDGNFL--LVMAAYDRYVAICFPLHYMSIMPKI-ZVSLVAVLSVL 123
QY 109 PVRFHMVSTVAAQSLCFLGLISLSTLVDPPLSYSHFVLEKFGNKLTVQMTDMIKS 168
DB 124 -TFH-----AMLHTLIMARLSFCEDSVIPHYCDMSTLKVACSDTHDN 167
QY 169 LTLAIAIGSPILYL-FL-----KIFDKFPDPLWYIMVFLVQI 207
DB 168 ELAIFILGPIYVLPFLLIIVSYARIYSSIFKVPSSQSIHKRFSGISLVSFLYGTV 227
QY 208 IAM-----TIIPVIM-PMFNKFT-PLEDELKKSIESLADRVGFP 246
DB 228 IGLVLCPSANNSEVKETWMSIYTVPMLNPFYISLRNDIKDALIKMKKQIP 281

RESULT 10
PCT-US93-08528-67
Sequence 67, Application PC/TUS9308528
GENERAL INFORMATION:

APPLICANT: New York University
TITLE OF INVENTION: POLYPEPTIDES OF G-COUPLED PROTEIN
NUMBER OF SEQUENCES: 348
CORRESPONDENCE ADDRESS:
ADDRESSEE: BROWDY AND NEIMARK
STREET: 419 Seventh Street, N.W., Suite 300
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US93/08528
FILING DATE: 09-SEP-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/943,236
FILING DATE: 10-SEP-1992
ATTORNEY/AGENT INFORMATION:
NAME: Townsend, Kevin G.

Tue Feb 12 16:22:25 2002

us-09-165-460a-2.rai

Page 7

```

?      REGISTRATION NUMBER: 34,033
?      REFERENCE/DOCKET NUMBER: MURPHY=2 PCT
?      TELECOMMUNICATION INFORMATION:
?      TELEPHONE: 202-628-5197
?      TELEFAX: 202-737-3528
?      TELEX: 246633
?      INFORMATION FOR SEQ ID NO: 67:
?      SEQUENCE CHARACTERISTICS:
?      LENGTH: 284 amino acids
?      TYPE: amino acid
?      STRANDEDNESS: single
?      TOPOLOGY: linear
?      MOLECULE TYPE: peptide
?      PCT-0593-08528-67

```

| | | | | |
|------------------------|------------------|----------------|------------|-------------|
| Query Match Similarity | 3.9% | Score 92.5; | DB 5; | Length 284; |
| Best Local Similarity | 21.1% | Pred. No. 0 | 14; | |
| Matches 62; | Conservative 45; | Mismatches 96; | Indels 91; | Gaps 14; |

[illegible]

RESULT 11
US-09-091-117-2
Sequence 2, Application US/09091117
Patent No. 6171589
GENERAL INFORMATION:
APPLICANT: The University of Melbourne
TITLE OF INVENTION: Mycoplasma Recombinant Polypeptides and
TITLE OF INVENTION: Vaccines
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: GREENEER, WINNER and SULLIVAN P. C.
STREET: 5370 Manhattan Circle, Suite 201
CITY: Boulder
STATE: Colorado
COUNTRY: United States of America
ZIP: 80303
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/091,117
FILING DATE: 12 JUNE 1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/AU96/00803
FILING DATE: 13-DEC-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: AU PNT127
FILING DATE: 13-DEC-1995
ATTORNEY/AGENT INFORMATION:
NAME: WINNER, Ellen P.

TELECOMMUNICATION INFORMATION
TELEPHONE: +1 303 499 8080
TELEFAX: +1 303 499 8089
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1030 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-091-117-2

| | | | | |
|---------------------------|-------|----------------|------------|-------------|
| Query Match | 3.9% | Score 92.5 | DB 4 | Length 1030 |
| Best Local Similarity | 20.9% | Pred. No. 0.94 | | |
| Matches 116, Conservative | 88 | Mismatches 181 | Indels 169 | Gaps 31 |

[illegible]

RESULT 12
 US-07-582-945-2
 : Sequence 2, Application US/07582945
 : Patent No. 5369019
 :
 : GENERAL INFORMATION:
 :
 : APPLICANT: TAKKER FOGED, Niels
 :
 : APPLICANT: PETERSON, Svend
 :
 : TITLE OF INVENTION: A PASTEURILLA VACCINE
 :
 : NUMBER OF SEQUENCES: 2
 :
 : CORRESPONDENCE ADDRESS:
 : ADDRESSSEE: Foley & Lardner
 : STREET: 1800 Diagonal Road, Suite 500
 : CITY: Alexandria
 : STATE: VA
 :
 : COUNTRY: USA
 :
 : ZIP: 22313-0299
 :

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA: US/07/582.945
APPLICATION NUMBER: US/07/582.945
FILING DATE: 19901012
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 30307/112 PLVI
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703)836-9300
TELEFAX: (703)683-4109
TELEX: 899149
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1285 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-582-945-2

Query Match 3.9% Score 92; DB 1; Length 1285;
Best Local Similarity 19.5% Pred. No. 1.5; Mismatches 133; Indels 126; Gaps 19;
Matches 77; Conservative 59;

QY 110 VRFHNVSTVAGSLCFLGLSLSTLVDPPLSYSHFVEEKGFKLTVQMTIMKSL 169
DB 714 IRFALQDVI-SAPRGVAGAIPEAIDPVRK-----VIED----- 748
QY 170 TLAVAIGGPIILFLKIDKFPDPLWYIMVFLVQIILMTIIVFIMP-----MKNKF 224
DB 749 -----ISVFDKIOTNY-WELPAYESWNEGSNSRLPGLRESOSKGMLSKC 793
QY 225 TPLEDG-----ELKKSIESLADRVGFPDLKIFVIDGSKRSSHNAVFTGLPFTSKRI 276
DB 794 RIENSIVYIGHSEEMFISISPYNOVGPEYL-----YPTFFSM 834
QY 277 V-----LFDTLVNSN-STDEITAVLAHEIGHWOKNHIIVNVIFISQILHT 318
DB 835 LOEVGDLGFEGAFATRNPFNTLVSDRLSLMENTMLTFESFDYTPWDAIYGDINDE--Q 892
QY 319 FLIFSLFTSI-----YRTSFYN--TFGFLEKSTGSPVDPVITKEPIIITGMLFND 369
DB 893 FAAMINERIEKCMNTYRGVAFONSSKSIDFPLNMLT-TFIDNGLTE-----IAISD 943
QY 370 LITPLECAMQFVMSLISRTHEYQA-DAVAKKL-----GYKONLCRALIDLOIK-----NL 418
DB 944 L--PYDIOQELISQFLOGSNEKKTIDAMLFNLDKGDINGARFKLQSAKDNNIKRAIGH 1001
QY 419 STMNVDP---YSSYHYSHPTLAERSTALDYSEK 450
DB 1002 SDNSVFPFNPNPYKSLYKGNITAEIIEKLDREGOK 1036

RESULT 13
US-08-453-141-2
Sequence 2, Application US/08453141
Patent No. 5885589
GENERAL INFORMATION:
APPLICANT: FOGED, Niels T.
APPLICANT: PETERSEN, Svend
TITLE OF INVENTION: PASTEURELLA VACCINE
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESS: Foley & Lardner
STREET: 3000 K Street, N.W., Suite 500
CITY: Washington, D.C.
COUNTRY: USA

ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA: US/08/453.141
APPLICATION NUMBER: US/08/453.141
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/293,314
FILING DATE: 22-AUG-1994
APPLICATION NUMBER: US 07/582.945
FILING DATE: 12-OCT-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/DK89/00084
FILING DATE: 11-APR-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: DK 1995/88
FILING DATE: 04-DEC-1988
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 40389/102/ARKZO
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202)672-5300
TELEFAX: (202)672-5399
TELEX: 904136
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1285 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-453-141-2

Query Match 3.9% Score 92; DB 2; Length 1285;
Best Local Similarity 19.5% Pred. No. 1.5; Mismatches 133; Indels 126; Gaps 19;
Matches 77; Conservative 59;

QY 110 VRFHNVSTVAGSLCFLGLSLSTLVDPPLSYSHFVEEKGFKLTVQMTIMKSL 169
DB 714 IRFALQDVI-SAPRGVAGAIPEAIDPVRK-----VIED----- 748
QY 170 TLAVAIGGPIILFLKIDKFPDPLWYIMVFLVQIILMTIIVFIMP-----MKNKF 224
DB 749 -----ISVFDKIOTNY-WELPAYESWNEGSNSRLPGLRESOSKGMLSKC 793
QY 225 TPLEDG-----ELKKSIESLADRVGFPDLKIFVIDGSKRSSHNAVFTGLPFTSKRI 276
DB 794 RIENSIVYIGHSEEMFISISPYNOVGPEYL-----YPTFFSM 834
QY 277 V-----LFDTLVNSN-STDEITAVLAHEIGHWOKNHIIVNVIFISQILHT 318
DB 835 LOEVGDLGFEGAFATRNPFNTLVSDRLSLMENTMLTFESFDYTPWDAIYGDINDE--Q 892
QY 319 FLIFSLFTSI-----YRTSFYN--TFGFLEKSTGSPVDPVITKEPIIITGMLFND 369
DB 893 FAAMINERIEKCMNTYRGVAFONSSKSIDFPLNMLT-TFIDNGLTE-----IAISD 943
QY 370 LITPLECAMQFVMSLISRTHEYQA-DAVAKKL-----GYKONLCRALIDLOIK-----NL 418
DB 944 L--PYDIOQELISQFLOGSNEKKTIDAMLFNLDKGDINGARFKLQSAKDNNIKRAIGH 1001
QY 419 STMNVDP---YSSYHYSHPTLAERSTALDYSEK 450
DB 1002 SDNSVFPFNPNPYKSLYKGNITAEIIEKLDREGOK 1036

RESULT 14
US-08-293-314-2

Sequence 2, Application US/08293314
Patent No. 6110470
GENERAL INFORMATION:
APPLICANT: FOGED, Niels T.
APPLICANT: PETERSEN, Svend
TITLE OF INVENTION: PASTEURELLA VACCINE
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 3000 K Street, N.W., Suite 500
CITY: Washington, D.C.
COUNTRY: USA
ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/293,314
FILING DATE: 22-AUG-1994
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/582,945
FILING DATE: 12-OCT-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/DK89/00084
FILING DATE: 11-APR-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: DK 1995/88
FILING DATE: 04-DEC-1988
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 40389/102/ANZO
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202)672-5300
TELEFAX: (202)672-5399
TELEX: 904136
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1285 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-293-314-2

Query Match 3.9%; Score 92; DB 3; Length 1285;
Best Local Similarity 19.5%; Pred. No. 1.5; Indels 126; Gaps 19;
Matches 77; Conservative 59; Mismatches 133;

QY 110 VRFHWSTVAOSICFGLSLSTLVLDPLSYSHFVLEKEGFKLVQVLTMDIKSL 169
DB 714 INFALQDPTVTSAPFRGVGAIPDAIDFVKY-----VID----- 748
QY 170 TLAVAGCPILYFLKIDKPTDFLWYVFLVVOILAMTIPVIMP-----WENKF 224
DB 749 -----ISVPEKIQTNV-WELPAYESWNEGNSRLPGLRESQSKMLSKC 793
QY 225 TPLEDG-----ELKKSIESLADRVGFPLDKIFVIDSKRSSHSNAFTGLPTFSKRI 276
DB 794 RLTEMSLYIGHSEYEFYSISPYSNQVGPYEL-----YPTFFFSM 834
QY 277 V-----LFDTLVNSN-STDEITAVLAHEIGHMOKNHIYNAVIFSQLHT 318
DB 835 LQEVGDGEGEAFATRNENFTLVSDRLSLAMENTMLTSPETTPMDAIGYINDE--Q 892
QY 319 FLIFSLSFSI-----YRNTSFYN--TFGFLEKSTGSEFVDPVITKEPPIIGMLFND 369
DB 893 FAAMSINERIEKCMNTYRGVAFONSKSIDFLNNLT--TFINDGLTE-----IAISD 943
QY 370 LITPLECANQFVMSLISRHEVQ-DAVAKKL-----CYKONLCRALDLQIK-----NL 418

QY 944 L--PYDIYOQELISOFLOGSENKTKTDAMLFNLDKGDINGCAFKLQSAKDNKIRARIGH 1001
DB 419 STWNVDPVLT--YSSYHYSHPTLAERSTALDYSEK 450
DB 1002 SDNSVPPFNPNPKSLYXKGNIIAEMAIERKIDREGOK 1036

RESULT 15
US-07-991-867B-42
Sequence 42, Application US/07991867B
Patent No. 5476781
GENERAL INFORMATION:
APPLICANT: Moyer, Richard W.
APPLICANT: Hall, Richard L.
APPLICANT: Gruidl, Michael E.
TITLE OF INVENTION: No. 5476781el Entomopoxvirus Expression System
NUMBER OF SEQUENCES: 66
CORRESPONDENCE ADDRESS:
ADDRESSEE: David R. Saliwanchik
STREET: 2421 N.W. 41st Street, Suite A-1
CITY: Gainesville
STATE: FL
COUNTRY: USA
ZIP: 32606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/991,867B
FILING DATE: 12-DEC-1992
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO 92/14818
FILING DATE: 12-FEB-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/827,685
FILING DATE: 30-JAN-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/657,584
FILING DATE: 19-FEB-1991
ATTORNEY/AGENT INFORMATION:
NAME: Saliwanchik, David R.
REGISTRATION NUMBER: 31,794
REFERENCE/DOCKET NUMBER: UF114.C3
TELECOMMUNICATION INFORMATION:
TELEPHONE: 904-375-8100
TELEFAX: 904-372-5800
INFORMATION FOR SEQ ID NO: 42:
SEQUENCE CHARACTERISTICS:
LENGTH: 485 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-991-867B-42

Query Match 3.8%; Score 90; DB 1; Length 485;
Best Local Similarity 19.5%; Pred. No. 0.56;
Matches 62; Conservative 50; Mismatches 132; Indels 74; Gaps 14;

QY 104 LNAVLPVRHWSTVAOSICFGLSLSTLVLDPLSYSHFVLEKEGFKLVQVLT--W 161
DB 12 INVALDKQHLNKNKIMNNI-----TNEITLQDYQVSKIFGLKLNLSMLFW 61
QY 162 ITDMKSLTLAVAGCPILYFLKIDKPTDFLWYVFLVVOILAMTIPVIMPWF 221
DB 62 DTGMCKTTLAVYII-----KYIK--ELFPR--WIIIFIKK-----SLYIDPWL 101
QY 222 NKFTPLEDELKKSIESLADRVGFPLDKIFVIDSKRSSHSNAY---FTGLPTFSKRIYL 278

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 12, 2002, 13:25:22 ; Search time 277.71 Seconds

(Without alignments)
9100.842 Million cell updates/sec

Title: US-09-165-460a-3

Perfect score: 2948

Sequence: 1 tgaactgttgatgaacaaga.....ggggagagataaagaatcaca 2948

Scoring table:

IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 930621 segs, 428662619 residues

Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

N.Geneseq_1101.*
1: /SIDS2/gcgdata/geneseq/geneseqn/NA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/geneseqn/NA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseqn/NA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseqn/NA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseqn/NA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseqn/NA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseqn/NA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseqn/NA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseqn/NA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseqn/NA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseqn/NA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseqn/NA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseqn/NA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseqn/NA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseqn/NA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseqn/NA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseqn/NA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseqn/NA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseqn/NA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseqn/NA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseqn/NA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseqn/NA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB | ID | Description |
|------------|--------|-------------|--------|----|----------|--------------------|
| 1 | 2711.2 | 92.0 | 2850 | 19 | AAV17602 | Yeast RCE1 gene en |
| 2 | 175.2 | 5.9 | 936 | 22 | AAFS8252 | Oligonucleotide D1 |
| 3 | 175.2 | 5.9 | 936 | 22 | AAFS8254 | Oligonucleotide D1 |
| 4 | 175.2 | 5.9 | 936 | 22 | AAFS8257 | Oligonucleotide D1 |
| 5 | 175.2 | 5.9 | 936 | 22 | AAFS8259 | Oligonucleotide D2 |
| 6 | 175.2 | 5.9 | 936 | 22 | AAFS8262 | Oligonucleotide D2 |
| 7 | 175.2 | 5.9 | 936 | 22 | AAFS8255 | Oligonucleotide D1 |
| 8 | 174.8 | 5.9 | 936 | 22 | AAFS8252 | Oligonucleotide D1 |
| 9 | 174.8 | 5.9 | 936 | 22 | AAFS8254 | Oligonucleotide D1 |
| 10 | 174.8 | 5.9 | 936 | 22 | AAFS8257 | Oligonucleotide D1 |
| 11 | 174.8 | 5.9 | 936 | 22 | AAFS8259 | Oligonucleotide D2 |

| | | | | | | |
|----|-------|-----|--------|----|-----------|--------------------|
| 12 | 174.8 | 5.9 | 936 | 22 | AAFS8262 | Oligonucleotide D2 |
| 13 | 174.8 | 5.9 | 938 | 22 | AAFS8255 | Oligonucleotide D1 |
| 14 | 65.6 | 2.2 | 244 | 22 | AAFS8238 | Oligonucleotide D1 |
| 15 | 63.6 | 2.2 | 244 | 22 | AAFS8238 | Oligonucleotide D1 |
| 16 | 53.6 | 1.8 | 6265 | 20 | AAAX08523 | NBP46 (root lectin |
| 17 | 52.6 | 1.8 | 396 | 22 | AAAF94862 | Human ovarian can |
| 18 | 50.4 | 1.7 | 580073 | 18 | AAAF58840 | Mycoplasma genital |
| 19 | 50.2 | 1.7 | 612 | 22 | AAH71471 | Human cervical can |
| 20 | 49.8 | 1.7 | 2229 | 21 | AAAZ3425 | CDNA encoding huma |
| 21 | 49.6 | 1.7 | 268 | 22 | AAH70080 | Human cervical can |
| 22 | 49.4 | 1.7 | 309 | 22 | AAH70139 | Human cervical can |
| 23 | 49.2 | 1.7 | 2771 | 22 | AAH72860 | Human cervical can |
| 24 | 48.8 | 1.7 | 310 | 22 | AAH71505 | Human cervical can |
| 25 | 48.2 | 1.6 | 2046 | 21 | AAAG95472 | Human secreted pro |
| 26 | 48.2 | 1.6 | 6243 | 20 | AAAG09598 | Clostridium specie |
| 27 | 48 | 1.6 | 671 | 22 | AAAG05318 | Human secreted pro |
| 28 | 48 | 1.6 | 19124 | 18 | AAAT2882 | Plasmodium var-7 g |
| 29 | 48 | 1.6 | 19124 | 21 | AAAG2887 | Plasmodium var-7 p |
| 30 | 47.8 | 1.6 | 1117 | 21 | AAAG59475 | Nucleotide sequenc |
| 31 | 47.6 | 1.6 | 4185 | 22 | AAAG03730 | P. falciparum telo |
| 32 | 47.4 | 1.6 | 4237 | 19 | AAAG1487 | Human secreted pro |
| 33 | 47.2 | 1.6 | 2427 | 11 | AAAG04107 | Human pro-urokinas |
| 34 | 47 | 1.6 | 2447 | 19 | AAAG54587 | Human secretory pr |
| 35 | 47 | 1.6 | 2447 | 20 | AAAG25607 | Human secreted pro |
| 36 | 46.8 | 1.6 | 612 | 22 | AAH71471 | Human cervical can |
| 37 | 46.6 | 1.6 | 597 | 22 | AAH71472 | Human cervical can |
| 38 | 46.4 | 1.6 | 731 | 13 | AAAG31633 | Rsal restriction f |
| 39 | 46.4 | 1.6 | 2486 | 21 | AAAG97037 | Nucleotide sequenc |
| 40 | 46.4 | 1.6 | 5139 | 21 | AAAG70139 | Plasmodium falcipa |
| 41 | 46 | 1.6 | 240 | 18 | AAAT6782 | Staphylococcus aur |
| 42 | 45.8 | 1.6 | 621 | 22 | AAH71551 | Human cervical can |
| 43 | 45.8 | 1.6 | 1254 | 22 | AAAG05400 | Human secreted pro |
| 44 | 45.8 | 1.6 | 6033 | 21 | AAAG70152 | Plasmodium falcipa |
| 45 | 45.8 | 1.6 | 9789 | 17 | AAAT41852 | CDNA encoding Plas |

ALIGNMENTS

RESULT 1
AAV17602 standard; DNA; 2850 BP.
ID AAV17602:
XX
AC AAV17602:
XX
DT 20-JUL-1998 (first entry)
XX
DE Yeast RCE1 gene encoding Rce1p protein.
XX
KW RCE1 gene; Rce1p; a-factor convertase; GMAX protease;
KW zinc metalloprotease; yeast; prenylation; Ras; inhibitor; cancer;
KW colorectal carcinoma; pancreas carcinoma; leukaemia; therapy; ss.
XX
OS Saccharomyces cerevisiae.
XX
PN W09805786-A2.
XX
PD 12-FEB-1998.
XX
PF 06-AUG-1997; 97MO-US14777.
XX
PR 30-JUL-1997; 97US-0902774.
XX
PR 07-AUG-1996; 96US-0023491.
XX
PA (REGC) UNIV CALIFORNIA.
XX
PI Ashby MN, Boyartchuk VL, Rine JD;
DR WPI; 1996-145621/13.
PT Vector comprising nucleic acid coding for Afc1p and Rce1p proteins
PT involved in protein prenylation - useful for screening for
PT inhibitors of proteins, especially of mutated ras proteins involved

PT In cancerous conditions

XX Claim 11; Page 48-50; 59pp; English.

XX This DNA sequence comprises the yeast RCE1 (Ras and a-factor
CC convertase) gene that codes for a novel protein (see AAM48302),
CC designated Rce1p, that mediates the removal of AAX tripeptide from
CC CAAX proteins such as Ras following prenylation. RCE1 was isolated
CC using an autocrine arrest, sensitised selection for CAAX proteases.
CC Another CAAX protease gene, APCI (a-factor convertase) (see AAV17601)
CC was similarly identified. Vectors comprising the APCI or RCE1
CC gene, polypeptides encoded by such vectors and recombinant cells
CC transduced with the vectors are claimed. The novel Apc1p and Rce1p
CC (see AAM48302) proteins can be used to screen for inhibitors of
CC prenylation which can be used in the treatment of Ras dependent
CC cancers such as colorectal and exocrine pancreatic carcinomas or
CC myeloid leukemia.

CC (N.B. the translated sequence of this RCE1 gene corresponds to the
CC Rce1p amino acid sequence provided in the specification only in the
CC N-terminal region (amino acid residues 1-233)).

CC Sequence 2850 BP; 903 A; 504 C; 518 G; 925 T; 0 other;

Query Match 92.0%; Score 2711.2; DB 19; Length 2850;
Best Local Similarity 96.4%; Pred. No. 0;
Matches 2845; Conservative 0; Mismatches 3; Indels 102; Gaps 3;

OY 1 tgaactgttgaacaaagaagcctgaacagcatcaaaagcttgcagatgttcca 60
DB 1 tgaactgttgaacaaagaagcctgaacagcatcaaaagcttgcagatgttcca 60
OY 61 ttcaaaaattgtgaatacgaacaaagttcaacaagaatttcttgatnaatgaatt 120
DB 61 ttcaaaaattgtgaatacgaacaaagttcaacaagaatttcttgatnaatgaatt 120
OY 121 atataatgaatagggagacgctgaagcgaagaagaaattacagaatttatcac 180
DB 121 atataatgaatagggagacgctgaagcgaagaagaaattacagaatttatcac 180
OY 181 ttccagatgttgaacgatacgaagaagatttgaacttgaaagttagttccaaact 240
DB 181 ttccagatgttgaacgatacgaagaagatttgaacttgaaagttagttccaaact 240
OY 241 atccaacattttaaanaaacagcgcagtaacccaaacaacccaacgaacgacat 300
DB 241 atccaacattttaaanaaacagcgcagtaacccaaacaacccaacgaacgacat 300
OY 301 tgaagatcaatcgcacacagtaacgtagttcatcgcgcgcagttcagcgtcatt 360
DB 301 tgaagatcaatcgcacacagtaacgtagttcatcgcgcgcagttcagcgtcatt 360
OY 361 aagcccgtaactcttcacccaagagtagttatttctactagcggtagctctcac 420
DB 361 aagcccgtaactcttcacccaagagtagttatttctactagcggtagctctcac 420
OY 421 ttcccttaaatcagacagcagatgttcataaagtgttgagttgctctcttaacgta 480
DB 421 ttcccttaaatcagacagcagatgttcataaagtgttgagttgctctcttaacgta 480
OY 481 caaacgagacataatgtgaacttgaaatttaataaagagacataaagaaacttaattcc 540
DB 481 caaacgagacataatgtgaacttgaaatttaataaagagacataaagaaacttaattcc 540
OY 541 aagtttgaagaagcgtcctatgttgaatttgaatttgaatttgaatttgaatttga 600
DB 541 aagtttgaagaagcgtcctatgttgaatttgaatttgaatttgaatttgaatttga 600
OY 601 aaacacattgagtagcgaagaattgataccatttctgttgaggaagagaaata 660
DB 601 aaacacattgagtagcgaagaattgataccatttctgttgaggaagagaaata 660
OY 661 aaaaaacattgaaaaaatctcgttacttctctatagatatagatatagtagtgttt 720

DB 661 aaaaaacattgaaaaaatctcgttacttctctatagatatagatatagtagtgttt 720
OY 721 gcttataagaagattatcgcgtcccttgatctcccatattataaattctt 780
DB 721 gcttataagaagattatcgcgtcccttgatctcccatattataaattctt 780
OY 781 taaatgattctcgggcctttgtgtcctgatttttttttttttttttttttttttttttt 840
DB 781 taaatgattctcgggcctttgtgtcctgatttttttttttttttttttttttttttttt 840
OY 841 gatgaaaaaccttgatgtatttataccttattttaagttactaaataacgcagatt 900
DB 841 gatgaaaaaccttgatgtatttataccttattttaagttactaaataacgcagatt 900
OY 901 cagagacaacacatagaatttcttcttgcaagaanaataaagaanaatttgatgctc 960
DB 901 cagagacaacacatagaatttcttcttgcaagaanaataaagaanaatttgatgctc 960
OY 961 tgactactgactgtctgcataagagaagacagaacagcaatgctacaaattcacaact 1020
DB 961 tgactactgactgtctgcataagagaagacagaacagcaatgctacaaattcacaact 1020
OY 1021 tctagtgctcctatacatctccatctcctatgtgctacgcgtatatagcaactcacaac 1080
DB 1021 tctagtgctcctatacatctccatctcctatgtgctacgcgtatatagcaactcacaac 1080
OY 1081 agaagggcttaacagagataatcccgagaagataaattctgcagacaaaacttaacat 1140
DB 1081 agaagggcttaacagagataatcccgagaagataaattctgcagacaaaacttaacat 1140
OY 1141 tatgtaattccaaacctttttgtgctcctttttaaactcctaattatctagtaacac 1200
DB 1141 tatgtaattccaaacctttttgtgctcctttttaaactcctaattatctagtaacac 1200
OY 1201 ttcaacataaagttcagaagacgaatttttaggttagtatacccaaggttcaacgc 1260
DB 1201 ttcaacataaagttcagaagacgaatttttaggttagtatacccaaggttcaacgc 1260
OY 1261 tgcattgcacaaaccttgagcaattcagcagatctgtgaagaacttaagaaatgtgtgc 1320
DB 1261 tgcattgcacaaaccttgagcaattcagcagatctgtgaagaacttaagaaatgtgtgc 1320
OY 1321 gatgtattgacccctatatttggaacggttttagatttttatatacattatataa 1380
DB 1321 gatgtattgacccctatatttggaacggttttagatttttatatacattatataa 1380
OY 1381 tccaaagagctctacttgagaatttttaacagaaattcctgaatttttgagtttcaag 1440
DB 1381 tccaaagagctctacttgagaatttttaacagaaattcctgaatttttgagtttcaag 1440
OY 1441 gaatttataatttgcacacaataactgaggaataatttacaacgccaatgcttgactac 1500
DB 1441 gaatttataatttgcacacaataactgaggaataatttacaacgccaatgcttgactac 1500
OY 1501 gtacttaaaccttaacgcagcttcgcaactgaagatcaacaagttattttggcaacatc 1560
DB 1501 gtacttaaaccttaacgcagcttcgcaactgaagatcaacaagttattttggcaacatc 1560
OY 1561 gcttttttttgacttgcgcgcacacacagcttatgagaattacagaagaagctcat 1620
DB 1561 gcttttttttgacttgcgcgcacacacagcttatgagaattacagaagaagctcat 1620
OY 1621 gacaacgttctcactcgtcgcacaacatgcttccaatttatacaacaacattttg 1680
DB 1621 gacaacgttctcactcgtcgcacaacatgcttccaatttatacaacaacattttg 1680
OY 1681 aggtttaacaagtttgaattcgaagaacagcggaactaagtgctgcaatacct 1740
DB 1681 aggtttaacaagtttgaattcgaagaacagcggaactaagtgctgcaatacct 1740
OY 1741 gcatgcctttgcaataatcagtggttccctggtccttcaagaattgaatttcaattcac 1800

```

Db 1700 -----ttgcaatacatggygtttaatggtccttcaagattgatacttcaac 1750
QY 1801 agtagtagacaagaacgtgacgacatttccaaattgg--tctaaactggaataatg 1858
Db 1751 agtagtagacaagaacgtgacgacatttccaaattggccttcaactggaataatg 1810
QY 1859 tactgcgactgctgctgctggaataatccctggaaggaataccttcaactgga 1918
Db 1811 tactgcgactgctgctgctggaataatccctggaaggaataccttcaactgga 1870
QY 1919 ggaactcctggttatagaataacccctttagccttctttagctgatacgtttaa 1978
Db 1871 ggaactcctggttatagaataacccctttagccttctttagctgatacgtttaa 1930
QY 1979 atttccatgtactataactcttcttcaactataatggaattctatcgagcgagg 2038
Db 1931 atttccatgtactataactcttcttcaactataatggaattctatcgagcgagg 1990
QY 2039 ctttgttgcggaagtggaataacgagttttgtggtttgttgaagaattgagg 2098
Db 1991 ctttgttgcggaagtggaataacgagttttgtggtttgttgaagaattgagg 2050
QY 2099 actataagtaactataactcttgttatacggactcaataacagtcgttgcgtgag 2158
Db 2051 actataagtaactataactcttgttatacggactcaataacagtcgttgcgtgag 2110
QY 2159 attgaagtttcagatacctaagaagtagagaagaagtgagcatcctaagtttcga 2218
Db 2111 attgaagtttcagatacctaagaagtagagaagaagtgagcatcctaagtttcga 2170
QY 2219 tcttttttaaggttttatttggctcctcagaatttaaggtttagtttttttt 2278
Db 2171 tcttttttaaggttttatttggctcctcagaatttaaggtttagtttttttt 2230
QY 2279 gttttgtggtttacataatttcaattcaaaaggagaattttagctgtcttataatg 2338
Db 2231 gttttgtggtttacataatttcaattcaaaaggagaattttagctgtcttataatg 2290
QY 2339 atagagataacgagagatctgctgtactacataacgagataagcgatcgctagtaa 2398
Db 2291 atagagataacgagagatctgctgtactacataacgagataagcgatcgctagtaa 2350
QY 2399 gggataacggaagtcctgaaggtttgttctgtcagacgtcccttgcacatgaatc 2458
Db 2351 gggataacggaagtcctgaaggtttgttctgtcagacgtcccttgcacatgaatc 2410
QY 2459 ggcgtaaaaaacctaaaaagattgtcgaattggttcaatggaatttactatgtccag 2518
Db 2411 ggcgtaaaaaacctaaaaagattgtcgaattggttcaatggaatttactatgtccag 2470
QY 2519 tagataaaaaatcgtgtgggaatctagtgggagagatcagttctgtgcagacatca 2578
Db 2471 tagataaaaaatcgtgtgggaatctagtgggagagatcagttctgtgcagacatca 2530
QY 2579 gtctgtgctcaatgccaagtgacacacccgtttaaataacggaatgagatccaa 2638
Db 2531 gtctgtgctcaatgccaagtgacacacccgtttaaataacggaatgagatccaa 2590
QY 2639 ctccgctagagaacttgcattgaggggttaactcagagatagaatccctccataag 2698
Db 2591 ctccgctagagaacttgcattgaggggttaactcagagatagaatccctccataag 2650
QY 2699 aacaggttaactactatgataaaaaaagggttcaactcctccacaggaatgaatg 2758
Db 2651 -----aaatlaatg 2660
QY 2759 ctaacgtattaagaagaaacttattatggttccgcgaatcaaacctcaagtttaag 2818
Db 2661 ctaacgtattaagaagaaacttattatggttccgcgaatcaaacctcaagtttaag 2720
QY 2819 ctgataatcttctagagctgtgtacaagatactttacaaaatatacaactaaagcaca 2878
Db 2721 ctgataatcttctagagctgtgtacaagatactttacaaaatatacaactaaagcaca 2780

```

```

QY 2879 gtgaagataatgatgggaataagcaatgaataaacgataattgaggaatgggagata 2938
Db 2781 gtgaagataatgatgggaataagcaatgaataaacgataattgaggaatgggagata 2840
QY 2939 aagaatcaca 2948
Db 2841 aagaatcaca 2850

RESULT 2
AAF58252/c
ID AAF58252 standard; DNA; 936 BP.
XX
AC AAF58252;
XX
DT 24-APR-2001 (first entry)
XX
DE Oligonucleotide D1835.
XX
KW Electron-transfer group; ETM; mismatch; genotyping;
KW gene expression; ss.
OS Synthetic.
XX
PN WO200107665-A2.
XX
PD 01-FEB-2001.
XX
PF 26-JUL-2000; 2000MO-US20476.
XX
PR 26-JUL-1999; 99US-0145695.
PR 17-MAR-2000; 2000US-0190259.
XX
PA (CLIN-) CLINICAL MICRO SENSORS INC.
XX
PI UmeK RM;
XX
DR WPI; 2001-159728/16.
XX
PT Nucleic acids containing electron-transfer group, useful as labels in
PT hybridization assays, e.g. for genotyping, allowing repeat analyses on
PT a single surface
XX
PS Example 6: Page 127; 159pp: English.
XX
CC The present invention relates to a composition comprising two nucleic
CC acids each containing an electron-transfer group (ETM) having
CC different redox potentials. The invention is used for electronic
CC detection of nucleic acids, especially of substitutions (mismatches)
CC and single-nucleotide polymorphisms, e.g. for genotyping,
CC monitoring gene expression.
XX
SQ Sequence 936 BP; 4 A; 139 C; 10 G; 7 T; 776 other;

Query Match 5.9%; Score 175.2; DB 22; Length 936;
Best Local Similarity 1.1%; Pred. No. 2,1e-29;
Matches 9; Conservative 529; Mismatches 252; Indels 0; Gaps 0;

QY 456 ggttagatggtccttaccagtaacagagacatgaatgtgaacttggaatttaata 515
Db 795 GGGGGGGCMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMM 736
QY 516 aggaactaaagaaacttaattccaaagtlttgaagaagcctagtgtagtttatt 575
Db 735 MWMWMWMWMWMWMWMWMWMWMWMWMWMWMWMWMWMWMWMWMWMWMWMWMWMWM 676
QY 576 ggcgtcagtaatgaatgaatttgaaaacacatcttgcgctgagagatgatcccta 635
Db 675 MWMWMWMWMWMWMWMWMWMWMWMWMWMWMWMWMWMWMWMWMWMWMWMWMWMWM 616
QY 636 ttctgttaggcaagtgcacaataataaaaacattagaanaaatctctgttacttctt 695

```

[illegible]

XX Nucleic acids containing electron-transfer group, useful as labels in
PT hybridization assays, e.g. for genotyping, allowing repeat analyses on
PT a single surface -
XX
XX
PS Example 6; Page 127; 159pp; English.
PS
XX
XX The present invention relates to a composition comprising two nucleic
CC acids each containing an electron-transfer group (ETM) having
CC different redox potentials. The invention is used for electronic
CC detection of nucleic acids, especially of substitutions (mismatches)
CC and single-nucleotide polymorphisms, e.g. for genotyping,
CC monitoring gene expression.
XX
XX
SQ Sequence 936 BP; 4 A; 144 C; 7 G; 5 T; 776 other;

| | | | | |
|-----------------------|-------|--------------------|-----------------|-------------|
| Query Match | 5.98; | Score 175.2; | DB 22; | Length 936; |
| Best Local Similarity | 1.18; | Pred. No. 2.1e-29; | | |
| Matches | 9; | Conservative 529; | Mismatches 252; | Indels 0; |
| | | | Gaps | 0; |

| | | | |
|----|------|--|------|
| QY | 456 | gggttagatctgccttcacgtacaacggaacatacttgaaacttgatgaattaa | 515 |
| Db | 795 | GGGGGGGCGC | 736 |
| QY | 516 | aggacattaagaaacttcaacttccaagtttggaaagctgcatactgttgattat | 575 |
| Db | 735 | W | 676 |
| QY | 576 | gcgttcgagtaagatgaattgaataaccacatctggcgtagcgaagattgatacccta | 635 |
| Db | 675 | W | 616 |
| QY | 636 | ttctcgttagcgaagtcacaataaaaaacatcagaanaatctcgttaacttctt | 695 |
| Db | 615 | W | 556 |
| QY | 696 | atagataagatatactgattcttgcttataagatgaattatcgcgtcctctga | 755 |
| Db | 555 | W | 496 |
| QY | 756 | ttcccatctataataaattctttaaaatgcatcttcctgctgctcttgctctc | 815 |
| Db | 495 | W | 436 |
| QY | 816 | gtatttttttttttggaccactgtagyaaaacttgatgattataacttaatt | 875 |
| Db | 435 | W | 376 |
| QY | 876 | ttaaagttaacaaatatacagatcttcaggaacaaacatagatacttcttgcaga | 935 |
| Db | 375 | W | 316 |
| QY | 936 | aataaaacgaataaattgactgttgaactgaactgctgctcatagagagaacga | 995 |
| Db | 315 | W | 256 |
| QY | 996 | cagaatgcataactctcaaaattctagtcgtcctataaactccacatatactatg | 1055 |
| Db | 255 | W | 196 |
| QY | 1056 | taccgcatactgaacttccaaacgaaggtctaaacagagataatctcgaacgata | 1115 |
| Db | 195 | W | 136 |
| QY | 1116 | aattcgcatacgaacaaactacaatactgctaattccaactttttgtgctcttt | 1175 |
| Db | 135 | W | 76 |
| QY | 1176 | tacaatctcaatatatactagtaacctcacaataaagtttcaagaagcattttaagct | 1235 |
| Db | 75 | W | 16 |

[illegible]

```

Query Match: 5.9%; Score 175.2; DB 22; Length 938;
Best Local Similarity 1.1%; Pred. No. 2.1e-29;
Matches 9; Conservative 529; Mismatches 252; Indels 0; Gaps

QY 456 gggttagatgtccttccttaacgacacaagagacataatgtgaacttggaattaa 515
    ||| |:: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: :::
Db 795 GGGGGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 736

QY 516 aggacattaggaacttaattccaaqgtttgaagcgccatgttttagtttatt 575
    ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: :::
Db 735 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 676

QY 576 ggcgttcgaatgatbataaattgaaacatcttgctgtagcgaagatgtaacccta 635
    ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: :::
Db 675 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 616

QY 636 ttctgttagcgaagtgcacaataaataacattgaaaaaattctctgaacttcct 695
    ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: :::
Db 615 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 556

QY 696 atagatatagatatgtatggtttgtctatagatgaagtatttatcgcgctcttga 755
    ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: :::
Db 555 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 496

QY 756 ttcccatltaataaattccttaaaatgcatcttcctggtcctttgtcctct 815
    ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: :::
Db 495 WGGWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 436

QY 816 gtatttttttttttttgaccactgtagtaaaccttgatgatatttaacttatt 875
    ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: :::
Db 435 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 376

QY 876 ttaagttactaaatatacgaatcttcagacaaaacatagaaatttctttgcaagaa 935
    ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: :::
Db 375 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 316

QY 936 aataaacaagaataaattgattgcttgactactgactgctgctatagagagaaccagaa 995
    ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: :::
Db 315 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 256

QY 996 cagaatgctacaattctcaacattctagtgctcctataaacttcacatactcctatg 1055
    ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: :::
Db 255 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 196

QY 1056 tacgcgtatatacgaacttcacacaacagaagtgctaaacagagataatcccgacagat 1115
    ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: :::
Db 195 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 136

QY 1116 aatctgcatacgaacaaacttacaaatctgctaatcttccaaacttttttggcgctttt 1175
    ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: :::
Db 135 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 76

QY 1176 tacaactcaatctacagtacccactcacataaaagtttccaagagcgcattttaagct 1235
    ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: :::
Db 75 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 16

QY 1236 taagtattat 1245
    ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: :::
Db 15 WWWWWWWWWWWWWWWW 6

RESULT 8
AAFS8252
ID AAF58252 standard; DNA; 936 BP.
XX
AC AAF58252;

```

```

XX 24-APR-2001 (first entry)
DT Oligonucleotide D1835.
DE
XX Electron-transfer group; ETM; mismatch; genotyping;
KW gene expression; ss.
XX Synthetic.
OS
XX WO200107665-A2.
XX
XX 01-FEB-2001.
PD
XX 26-JUL-2000; 2000WO-US20476.
PF
XX 26-JUL-1999; 99US-0145695.
PR 17-MAR-2000; 2000US-0190259.
XX
XX (CLIN-) CLINICAL MICRO SENSORS INC.
XX
XX Umek RM;
XX WPI: 2001-159728/16.
XX
XX Nucleic acids containing electron-transfer group, useful as labels in
PT hybridization assays, e.g. for genotyping, allowing repeat analyses on
PT a single surface.
XX
XX Example 6; Page 127; 159pp; English.
XX
XX The present invention relates to a composition comprising two nucleic
CC acids each containing an electron-transfer group (ETM) having
CC different redox potentials. The invention is used for electronic
CC detection of nucleic acids, especially of substitutions (mismatches)
CC and single-nucleotide polymorphisms, e.g. for genotyping.
CC monitoring gene expression.
XX
XX Sequence 936 BP; 4 A; 139 C; 10 G; 7 T; 776 other;
SQ
Query Match 5.9%; Score 174.8; DB 22; Length 936;
Best Local Similarity 0.6%; Pred. No. 2.6e-29;
Matches 5; Conservative 530; Mismatches 247; Indels 0; Gaps 0;
QY 450 ataaatgggttagagcttcctccttaacagtaacagagacattatgtaacttgat 509
Db 1 www. ....
QY 510 ttaataaggacataaaggaaacttaattccaagtttgaagaagctgctatgttagt 569
Db 61 www. ....
QY 570 tttaatgcgtcgaagtaagtaattgaataaccactctggcgtgagaagattgata 629
Db 121 www. ....
QY 630 tccctattctgttagagcaagtgacaaaataaaaacattagaanaaaattctgttact 689
Db 181 www. ....
QY 690 ttctctatagatagatagatagctgtctctatagatagaagatattatcgctcc 749
Db 241 www. ....
QY 750 ttctgattcctattatataaataattctttaaatgcatlctctgtgctcttgg 809
Db 301 www. ....
QY 810 gctcgtgatttttttttttttgacacctgagtgaaacaccttgatgatttatacc 869
Db 361 www. ....

```

```

QY 870 ttatttaagttactaaatatcgagatttcaggaacaaacatagatttcttggc 929
Db 421 www. ....
QY 930 aagaaaataaagcaataaattgagcttctgactgactgctgctcatagagaa 989
Db 481 www. ....
QY 990 ccagagacagcaatgcttaacattcacaattctagctctccttaacatcccatatct 1049
Db 541 www. ....
QY 1050 atgigtacgcgtatattgcaacttcacacaggaaggctcaacagataatcctgaa 1109
Db 601 www. ....
QY 1110 cgalttaactcgcattgcaaaaacttaacattatgcttaattccaaccttttggc 1169
Db 661 www. ....
QY 1170 ctttttaacatctcaattatctagtcaccacttcacatatagttcaagagcattt 1229
Db 721 www. ....
QY 1230 ta 1231
Db 781 ww 782

```

RESULT 9
AAF58254
ID AAF58254 standard; DNA; 936 BP.
XX
AC AAF58254;
XX
DT 24-APR-2001 (first entry)
XX
DE Oligonucleotide D1875.
XX
KW Electron-transfer group; ETM; mismatch; genotyping;
KW gene expression; ss.
XX
OS Synthetic.
XX
PN WO200107665-A2.
XX
PD 01-FEB-2001.
XX
PF 26-JUL-2000; 2000WO-US20476.
XX
PR 26-JUL-1999; 99US-0145695.
PR 17-MAR-2000; 2000US-0190259.
XX
PA (CLIN-) CLINICAL MICRO SENSORS INC.
XX
PI Umek RM;
XX
DR WPI: 2001-159728/16.
XX
PT Nucleic acids containing electron-transfer group, useful as labels in
PT hybridization assays, e.g. for genotyping, allowing repeat analyses on
PT a single surface.
XX
PS Example 6; Page 127; 159pp; English.
XX
CC The present invention relates to a composition comprising two nucleic
CC acids each containing an electron-transfer group (ETM) having
CC different redox potentials. The invention is used for electronic
CC detection of nucleic acids, especially of substitutions (mismatches)
CC and single-nucleotide polymorphisms, e.g. for genotyping,
CC monitoring gene expression.
XX
SQ Sequence 936 BP; 4 A; 144 C; 7 G; 5 T; 776 other;

Query Match 5.9%; Score 174.8; DB 22; Length 936;
 Best Local Similarity 0.6%; Pred. NO. 2.6e-29;
 Matches 5; Conservative 530; Mismatches 247; Indels 0; Gaps 0;

```

QY 450 ataaatgggttagatgcttccttaccagtaacagagacataatgtgaacttgaat 509
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1  www..... 60

QY 510 ttaataaggacataaagaaactttaatccaaglttgaagctgctaigtgtagt 569
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 61  www..... 120

QY 570 ttatcgctcgagtaatgaattgaaccatcttgcgttagcgaagtgtata 629
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 121 www..... 180

QY 630 tcccatctctgttagcaagtgcaaaaataaaaacattagaaaaatctcgttact 689
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 181 www..... 240

QY 690 ttcttataagataagataatgtatgcttgcctatagatgaaggtattatcgcgtcc 749
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 241 www..... 300

QY 750 ttgtatcccatattataaataatcctttaaatgcatcttctgctcttctt 809
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 301 www..... 360

QY 810 gcttcgtatcttcttcttcttgaccactgtatggaacaccttgaattatcacc 869
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 361 www..... 420

QY 870 ttattttaagttactaaataatcgaatttcaggaacaaacatagaaatttcttctg 929
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 421 www..... 480

QY 930 aagaaaaataaagaaataatgtatgcttgcctactgactgctgtctcatagagaa 989
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 481 www..... 540

QY 990 ccgaacagcaatgctacaattcacaattcagtgctcctatacatctccatctc 1049
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 541 www..... 600

QY 1050 atgtgtacgcgtatagcaactcacaacagaagggtctaacgagataatcctcgaa 1109
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 601 www..... 660

QY 1110 cgattaaatcgcgatacgaataacttacaattatgctaattccaactttttgtg 1169
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 661 www..... 720

QY 1170 ctttttacaatccaatctagtaacttccacataagtttcaagacgcatctt 1229
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 721 www..... 780

QY 1230 ta 1231
   : :
Db 781 ww 782

```

RESULT 10

AAFS8257 standard; DNA; 936 BP.

AAFS8257;

24-APR-2001 (first entry)

Oligonucleotide D1954.

XX

KW Electron-transfer group; ETM; mismatch; genotyping;
 KW gene expression; ss.

XX
 XX
 0S Synthetic.

XX WO200107665-A2.

XX PD 01-FEB-2001.

XX PE 26-JUL-2000; 2000MO-US20476.

XX PR 26-JUL-1999; 99US-0145695.

XX PR 17-MAR-2000; 2000US-0190259.

XX PA (CLIN-) CLINICAL MICRO SENSORS INC.

XX PI Uneak RM;

XX DR WPI; 2001-159728/16.

XX PT Nucleic acids containing electron-transfer group, useful as labels in hybridization assays, e.g. for genotyping, allowing repeat analyses on a single surface

XX PS Example 6; Page 127; 159pp; English.

XX CC The present invention relates to a composition comprising two nucleic acids each containing an electron-transfer group (ETM) having different redox potentials. The invention is used for electronic detection of nucleic acids, especially of substitutions (mismatches) and single-nucleotide polymorphisms, e.g. for genotyping, monitoring gene expression.

XX CC

XX XX

SO Sequence 936 BP; 5 A; 142 C; 7 G; 6 T; 776 other;

Query Match 5.9%; Score 174.8; DB 22; Length 936;
 Best Local Similarity 0.6%; Pred. NO. 2.6e-29;
 Matches 5; Conservative 530; Mismatches 247; Indels 0; Gaps 0;

```

QY 450 ataaatgggttagatgcttccttaccagtaacagagacataatgtgaacttgaat 509
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1  www..... 60

QY 510 ttaataaggacataaagaaactttaatccaaglttgaagctgctaigtgtagt 569
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 61  www..... 120

QY 570 ttatcgctcgagtaatgaattgaaccatcttgcgttagcgaagtgtata 629
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 121 www..... 180

QY 630 tcccatctctgttagcaagtgcaaaaataaaaacattagaaaaatctcgttact 689
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 181 www..... 240

QY 690 ttcttataagataagataatgtatgcttgcctatagatgaaggtattatcgcgtcc 749
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 241 www..... 300

QY 750 ttgtatcccatattataaataatcctttaaatgcatcttctgctcttctt 809
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 301 www..... 360

QY 810 gcttcgtatcttcttcttcttgaccactgtatggaacaccttgaattatcacc 869
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 361 www..... 420

QY 870 ttattttaagttactaaataatcgaatttcaggaacaaacatagaaatttcttctg 929
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 421 www..... 480

QY 930 aagaaaaataaagaaataatgtatgcttgcctactgactgctgtctcatagagaa 989

```

[illegible]

| | | | |
|--|------|--|------|
| Qy | 450 | ataaagtggttagagattgcccctttacagtaacaaagagatctaaatgtaactcggat | 509 |
| Db | 1 | www. | 60 |
| Qy | 510 | ttaaataggacatlaagaaacttlaattccaaagtttgaagctgtcattgttagt | 569 |
| Db | 61 | www. | 120 |
| Qy | 570 | tttatcgctcgagtaatgattaaatttgaaaaccatcttggcgtagcgaattgata | 629 |
| Db | 121 | www. | 180 |
| Qy | 630 | tcacattctgttagcgaatgacaaataaataaacattgaanaaatctcgttact | 689 |
| Db | 181 | www. | 240 |
| Qy | 690 | ttctctatagatagatagatagattgttgccttataagatgattatcgcgtcc | 749 |
| Db | 241 | www. | 300 |
| Qy | 750 | tttgtatccctatatataataatcttttaaatgcatcttctgtgtcctcttct | 809 |
| Db | 301 | www. | 360 |
| Qy | 810 | gctctgatcttttttttttggaccacgtagtgaacaacttgaatatttacc | 869 |
| Db | 361 | www. | 420 |
| Qy | 870 | ttatcttaagttaactaaatcgcgatttcaggaacaaacatagaatttcttctgc | 929 |
| Db | 421 | www. | 480 |
| Qy | 930 | aagaaaaataaacgaataatgtatgcttgcactcgcgtctgtctatagaagaa | 989 |
| Db | 481 | www. | 540 |
| Qy | 990 | ccagaacagcaatgcatactccaacttcagtgctcctatatactccatcct | 1049 |
| Db | 541 | www. | 600 |
| Qy | 1050 | atgtgctacgcgtatataatgcaactcgaacaggaaggtctaaacgagataactcctgaa | 1109 |
| Db | 601 | www. | 660 |
| Qy | 1110 | cgattaaatctgcatactgcaaaacttaacataatgctaattccaaacttttctgtgc | 1169 |
| Db | 661 | www. | 720 |
| Qy | 1170 | ctttcttaacatcacaattactagtaaccatcacataaagtcttcaagagcattt | 1229 |
| Db | 721 | www. | 780 |
| Qy | 1230 | ta 1231 | |
| Db | 781 | ww 782 | |
| RESULT 12 | | | |
| AAFS8262 | | | |
| ID AAF58262 standard; DNA; 936 Bp. | | | |
| AC AAF58262; | | | |
| XX | | | |
| DT 24-APR-2001 (first entry) | | | |
| XX | | | |
| DE Oligonucleotide D2007. | | | |
| XX | | | |
| KM Electron-transfer group; ETM; mismatch; genotyping; | | | |
| XX gene expression; ss. | | | |
| XX | | | |
| XX Synthetic. | | | |
| XX | | | |

| | | | |
|---|------|---|------|
| Oy | 510 | ttaataagaactaaaggaaccttaattccaaagttttgaagctgcgcatgttgtagt | 569 |
| Db | 61 | ##### | 120 |
| Oy | 570 | tttatcgctcggatgaatgaattaattgnaaacacatcttcgctagcgagaatgata | 629 |
| Db | 121 | ##### | 180 |
| Oy | 630 | lcccaatttcgttaggcgaagtgcacaataaaaaacattagaanaattctcgttact | 689 |
| Db | 181 | ##### | 240 |
| Oy | 690 | ttctctatagatatagatatgatgttgcttcctabagatgaaggattattaccgctcc | 749 |
| Db | 241 | ##### | 300 |
| Oy | 750 | tttgatgcccatatataataaatctttaaagcatcttcgctgctccttttgt | 809 |
| Db | 301 | ##### | 360 |
| Oy | 810 | gctctgatttttttttttggaccactgagatgaaaaccttgatatttattacc | 869 |
| Db | 361 | ##### | 420 |
| Oy | 870 | tttatctaagtctactaaatcgcgatttcaggaacaaacatagataattctctgtc | 929 |
| Db | 421 | ##### | 480 |
| Oy | 930 | aagaaataataacgaataatgatgcttgactactgactcgtctgcatagagaga | 989 |
| Db | 481 | ##### | 540 |
| Oy | 990 | cgaagacgcgaatgctacaaattccaacattctagtgctcctatalcatccatacct | 1049 |
| Db | 541 | ##### | 600 |
| Oy | 1050 | atgctgcacgcgtatgcaacttcacaccgaagggctctaaagagataactcotgaa | 1109 |
| Db | 601 | ##### | 660 |
| Oy | 1110 | cgattaaatcgcgcatgcaaaaacttacattatgcttaattccaaacctttttgtgc | 1169 |
| Db | 661 | ##### | 720 |
| Oy | 1170 | cttttttaacatcctaattactgataccactcacataagtttcaaagacgacttt | 1229 |
| Db | 721 | ##### | 780 |
| Oy | 1230 | ta 1231 | |
| Db | 781 | wv 782 | |
| RESULT_14 AAFS8238/C ID AAF58238 standard; DNA: 244 BP. | | | |
| XX | AC | AAF58238; | |
| XX | DT | 24-APR-2001 (first entry) | |
| XX | DE | Oligonucleotide D1250:D1102. | |
| XX | KW | Electron-transfer group; ETM; mismatch; genotyping; gene expression; ss. | |
| XX | OS | Synthetic. | |
| XX | PN | MO200107665-A2. | |
| XX | PD | 01-FEB-2001. | |
| XX | PF | 26-JUL-2000.2000MO-US20476. | |

```

xx      26-JUL-1999;    98US-0145695.
pr      17-MAR-2000; 2000US-0190259.
pa      (CLIN-) CLINICAL MICRO SENSORS INC.
xx      Umek RM;
xx      WPI: 2001-159728/16.
dr      Nucleic acids containing electron-transfer group, useful as labels in
pt      hybridization assays, e.g. for genotyping, allowing repeat analyses on
pr      a single surface -
xx      Example 4; Page 120; 159pp; English.
xx      The present invention relates to a composition comprising two nucleic
cc      acids each containing an electron-transfer group (ETM) having
cc      different redox potentials. The invention is used for electronic
cc      detection of nucleic acids, especially of substitutions (mismatches)
cc      and single-nucleotide polymorphisms, e.g. for genotyping,
cc      monitoring gene expression.
xx      Sequence 244 BP; 19 A; 9 C; 12 G; 10 T; 194 other;
so      .
Query Match          2.2%; Score 65.6; DB 22; Length 244;
Best Local Similarity 6.4%; Pred.No.2.1e-05;
Matches 14; Conservative 145; Mismatches 59; Indels 0; Gaps 0.
Qy      737 attatcgagcttcgttgcattccctactataataaatcctttaaaatgcatcttctg 796
Db      222 ATTGATGGTGTCTTTTACAAWWWWWMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMW 163
Qy      797 gtgccttttgtgctctgtaatttttttttttggaaccactgagtgaacaccttga 856
Db      162 WWWWWWMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMW 103
Qy      857 tgaatttatcccttatttaagttaactaaataatcgagattccaagacaacaacatag 916
Db      102 WWWWWWMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMW 43
Qy      917 aatttccttgcagaagaaaataaacgaataaactg 954
Db      42 WWWWWWMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMCTG 5
RESULT 15
AAF58238
AAF58238 standard; DNA; 244 BP.
XX      AAF58238;
AC      24-APR-2001 (first entry)
XX      24-APR-2001
DE      Oligonucleotide D1250:D1102.
XX      Electron transfer group; ETM; mismatch; genotyping;
KW      gene expression; ss.
XX      Synthetic.
OS      Synhtetic.
FN      WO200107665-A2.
PD      01-FEB-2001.
PE      26-JUL-2000; 2000WO-US20476.
PR      26-JUL-1999; 99US-0145695.
PA      17-MAR-2000; 2000US-0190259.
XX      (CLIN-) CLINICAL MICRO SENSORS INC.

```


THIS PAGE BLANK (LJPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 12, 2002, 12:01:41 ; Search time 2973.29 Seconds

(without alignments)
10634.378 Million cell updates/sec

Title: US-09-165-460A-3

Perfect score: 2948

Sequence: 1 tgaactgttgatgaacaaag.....ggggagagataaagatcaca 2948

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 11351937 segs, 5372889281 residues

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

EST: *
1: em_estfun: *
2: em_esthum: *
3: em_estlin: *
4: em_estom: *
5: em_estopl: *
6: em_estda: *
7: em_estro: *
8: em_estov: *
9: em_hic: *
10: gD_est1: *
11: gD_est2: *
12: gD_hic: *
13: gD_gss: *
14: em_gss_fun: *
15: em_gss_hum: *
16: em_gss_hiv: *
17: em_gss_pln: *
18: em_gss_pro: *
19: em_gss_rod: *
20: em_gss_vrt: *
21: em_gss_other: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|---------------------|
| 1 | 428 | 14.5 | 994 | 13 | CNS066GJ |
| 2 | 161.6 | 5.5 | 519 | 13 | AZ926837 17 end of |
| 3 | 140 | 4.7 | 1045 | 13 | AL408764 17 end of |
| 4 | 113 | 3.8 | 1134 | 13 | CNS06POM |
| 5 | 86.2 | 2.9 | 497 | 13 | AL406947 17 end of |
| 6 | 85.6 | 2.9 | 933 | 13 | AZ929776 479.dif88 |
| 7 | 83 | 2.8 | 1111 | 13 | CNS06PEO |
| 8 | 79.4 | 2.7 | 1101 | 13 | AL404740 17 end of |
| 9 | 75.6 | 2.6 | 1125 | 10 | CNS0039G |
| 10 | 75.2 | 2.6 | 1101 | 13 | AL347503 17 end of |
| 11 | 70.8 | 2.4 | 996 | 13 | AL069706 Drosophila |
| 12 | 66.8 | 2.3 | 1101 | 13 | CNS001FB |

| | | | | | | |
|------|------|-----|------|----|----------|---------------------|
| C 13 | 66.8 | 2.3 | 1101 | 13 | CNS003BD | AL064091 Drosophila |
| C 14 | 66.6 | 2.3 | 619 | 10 | AL514935 | AL514935 |
| C 15 | 66.4 | 2.3 | 840 | 13 | CNS0107W | AL099422 Drosophila |
| C 16 | 66 | 2.2 | 759 | 13 | CNS06QXV | AL411257 17 end of |
| C 17 | 65.4 | 2.2 | 1043 | 13 | CNS0145P | AL103735 Drosophila |
| C 18 | 65.4 | 2.2 | 1225 | 13 | CNS0161D | AL106171 Drosophila |
| C 19 | 64.6 | 2.2 | 404 | 10 | AL514087 | AL514087 |
| C 20 | 64.6 | 2.2 | 811 | 13 | AZ538053 | AZ538053 EMTGA06TF |
| C 21 | 64.4 | 2.2 | 1101 | 13 | CNS0127P | AL102007 Drosophila |
| C 22 | 64.2 | 2.2 | 329 | 10 | AL513719 | AL513719 |
| C 23 | 63.8 | 2.2 | 1101 | 13 | CNS00EFA | AL069119 Drosophila |
| C 24 | 63.4 | 2.2 | 955 | 13 | BH136744 | BH136744 EMTG02TF |
| C 25 | 63.4 | 2.2 | 1101 | 13 | CNS0039G | AL063921 Drosophila |
| C 26 | 63.4 | 2.2 | 1101 | 13 | CNS00EFA | AL069847 Drosophila |
| C 27 | 63 | 2.1 | 366 | 10 | AL514543 | AL514543 |
| C 28 | 63 | 2.1 | 1190 | 13 | CNS020M7 | AL20620M7 Tetradon |
| C 29 | 62.8 | 2.1 | 966 | 13 | CNS0055C | AL061991 Drosophila |
| C 30 | 62.6 | 2.1 | 335 | 10 | AL513597 | AL513597 |
| C 31 | 62.6 | 2.1 | 1001 | 13 | CNS0155H | AL175696 Tetradon |
| C 32 | 62.4 | 2.1 | 1101 | 13 | CNS00L72 | AL105023 Drosophila |
| C 33 | 62.2 | 2.1 | 1101 | 13 | CNS00K85 | AL078743 Drosophila |
| C 34 | 62 | 2.1 | 1092 | 13 | AL515191 | AL515191 |
| C 35 | 62 | 2.1 | 1092 | 13 | CNS020K7 | AL175696 Tetradon |
| C 36 | 62 | 2.1 | 1101 | 13 | CNS00PFG | AL071206 Drosophila |
| C 37 | 61.8 | 2.1 | 425 | 10 | AL514791 | AL514791 |
| C 38 | 61.8 | 2.1 | 1101 | 13 | CNS0021J | AL061936 Drosophila |
| C 39 | 61.6 | 2.1 | 748 | 13 | CNS03FED | AL241363 Tetradon |
| C 40 | 61.2 | 2.1 | 867 | 13 | CNS00CX5 | AL060052 Drosophila |
| C 41 | 61 | 2.1 | 625 | 13 | CNS036A2 | AL229763 Tetradon |
| C 42 | 61 | 2.1 | 625 | 13 | CNS036A2 | AL229763 Tetradon |
| C 43 | 60.8 | 2.1 | 758 | 10 | AV756150 | AV756150 |
| C 44 | 60.8 | 2.1 | 1101 | 13 | CNS00E07 | AL069440 Drosophila |
| C 45 | 60.4 | 2.0 | 353 | 10 | AL515235 | AL515235 |

ALIGNMENTS

| | | | | | |
|------------|--|-------------|-----|-----|-------------|
| RESULT 1 | CNS066GJ | 994 bp | DNA | GSS | 17-JUN-2001 |
| LOCUS | T7 end of clone AS0A002A09 of library AS0A from strain CLIB 533 | | | | |
| DEFINITION | of Saccharomyces bayanus, genomic survey sequence. | | | | |
| ACCESSION | AL397313 | | | | |
| VERSION | AL397313.1 | GI:12149936 | | | |
| KEYWORDS | GSS. | | | | |
| SOURCE | Saccharomyces bayanus. | | | | |
| ORGANISM | Saccharomyces bayanus | | | | |
| REFERENCE | Saccharomycetales; Saccharomycetaceae; Saccharomycetes. | | | | |
| AUTHORS | Elkaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes. | | | | |
| TITLE | 1 (bases 1 to 994) | | | | |
| JOURNAL | Bon, E., Nevegilise, C., Casaregola, S., Artiguenave, F., Wincker, P., | | | | |
| MEDLINE | Aigle, M. and Durrens, P. | | | | |
| REFERENCE | Genomic exploration of the hemiascomycetous yeasts: 5. | | | | |
| AUTHORS | Saccharomyces bayanus var. ivarium | | | | |
| TITLE | 2 (bases 1 to 994) | | | | |
| JOURNAL | Souciot, J.L., Aigle, M., Artiguenave, F., Blandin, G., | | | | |
| MEDLINE | de Montigny, J., Dujon, B., Durrens, P., Lepingle, A., Llorente, B., | | | | |
| REFERENCE | Malpertuy, A., Nevegilise, C., Ozler-Kalogeropoulos, O., Potier, S., | | | | |
| AUTHORS | Saurin, W., Tekala, F., Roffano-Nioche, C., Wesolowski-Louvel, M., | | | | |
| TITLE | Wincker, P. and Weissbach, J. | | | | |
| JOURNAL | Genomic exploration of the hemiascomycetous yeasts: 1. A set of | | | | |
| MEDLINE | yeast species for molecular evolution studies | | | | |
| REFERENCE | FEBS Lett. 487 (1), 3-12 (2000) | | | | |
| AUTHORS | 20584715 | | | | |
| TITLE | 3 (bases 1 to 994) | | | | |
| JOURNAL | Genoscope. | | | | |
| MEDLINE | Submitted (07-SEP-2000) Genoscope - Centre National de Sequencage, | | | | |
| REFERENCE | Submitted (07-SEP-2000) Genoscope - Centre National de Sequencage, | | | | |
| AUTHORS | Submitted (07-SEP-2000) Genoscope - Centre National de Sequencage, | | | | |
| TITLE | Submitted (07-SEP-2000) Genoscope - Centre National de Sequencage, | | | | |
| JOURNAL | Submitted (07-SEP-2000) Genoscope - Centre National de Sequencage, | | | | |

| | | | | |
|---|-----------------------------------|---|---|------|
| Oy | 1734 | | tactccgcagccttcgatcatcacatcgagggtttccggtgccccaagtgaatttac | 1793 |
| Dd | 90 | TGTTTGGCATTCCTTTTTGTATATATATATAGGCCTCCCGGTACATCAGACTCACGCTTGCC | 31 | |
| OY | 1794 | attcaacagtagtaga | 1809 | |
| Dd | 30 | ATTATTACCATGATTGA | 15 | |
| RESULT | 3 | | | |
| CNS06POM | | GSS | 05-JUL-2001 | |
| Locus | | | | |
| DEFINITION | | Sacharomyces exiguus, genomic survey sequence. | | |
| ACCESSION | | AJ408764 | | |
| VERSION | | AL408764.1 GI:12175649 | | |
| KEYWORDS | | GSS. | | |
| ORGANISM | | Sacharomyces exiguus . Sacharomycetes Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetidae; | | |
| REFERENCE | | Sacharomycetales; Saccharomycetaceae; Saccharomyces. 1 (bases 1 to 1045) Bon,E., Neuveglise,C., Lepingle,A., Wincker,P., Artiguenave,F., Galliardin,C. and Casaregola,S. Genomic exploration of the hemiascomycetous yeasts : 6. | | |
| TITLE | JOURNAL MEDLINE REFERENCE AITHORS | Sacharomyces exiguus FEBS Lett. 487 (1),42-46 (2000) | | |
| TITLE | JOURNAL MEDLINE REFERENCE AITHORS | 2 (bases 1 to 1045) Soucllet,J.L., Anglé,M., Artiguenave,F., Blandin,G., Boletin-Pukhareva,M., Bon.E., Brotlier.P., Casaregola,S., de-Montigny.J., Dujon,B., Durrens,P., Lepingle.A., Llorente.B., Melpertuy,A., Neuveglise.C., Ozier-Kalogeropoulos.O., Potter.S., Saurin,W., Tekala.F., Toffano-Nioche.C., Weslowski-Louvel.M., Wincker,P. and Weissenbach,J. Genomic exploration of the hemiascomycetous yeasts : 1. A set of yeast species for molecular evolution studies FEBS Lett. 487 (1), 3-12 (2000) | | |
| JOURNAL MEDLINE REFERENCE AITHORS JOURNAL TITLE COMMENT | | 3 (bases 1 to 1045) Genoscope. Direct Submission Submitted (07-SEP-2000) Genoscoope - Centre National de Sequençage, 2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail : segre@genoscope.cns.fr Web : www.genoscoope.cns.fr) This GSS is part of a random genomic sequencing program of thirteen yeast species: Sacharomyces bayanus var. uvarum, Saccharomyces exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii, Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila, Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to 5 kb were prepared and both extremities were sequenced. See keywords for description of this sequence and for the sequence of the other extremity of this insert. Location/Qualifiers 1..1045 /organism="Saccharomyces exiguus" /strain="CBS 379" /db_xref="taxon:34358" /cclone="AVOAA010E10" /clone_1lb="AVOAA" /note="end : T7" <128 ..>1042 /note-"similar to Saccharomyces cerevisiae ORF YMR275c [Bul1 ; ubiquitination pathway protein] " /evidence=not_experimental <134 ..>1042 /note-"similar to Saccharomyces cerevisiae ORF YM111w [strong similarity to ubiquitinlation protein Bul1p] " /evidence=not_experimental | | |
| misc_feature misc_feature misc_count | | 406 a 152 c 163 g 322 t 2 others | | |

| ORIGIN | Query Match | 4.7% | Score 140; | DB 13; | Length 1045; |
|------------|--|---|--------------------|-----------|--------------|
| | Best Local Similarity | 55.4%; | Pred. No. 2,5e-14; | | |
| | Matches 297; | Conservative 0; | Mismatches 230; | Indels 9; | Gaps 14; |
| QY | 65 | aaatctgtgataagaacacgaatccaagaatcttctgaaatlaagatlat | 124 | | |
| Db | 518 | AAAATCAAGATATATATTCAGATTCTACGAAATGCTGATTAATTCATCATTTAT | 577 | | |
| QY | 125 | aataatagaggagaccgttaggcacaaagagacacgaattacagatttacttca | 184 | | |
| Db | 578 | AATGTAATTCGTAGATTATTATTCCTGCGAATTAATAATTCACGATTCATCTAGT | 637 | | |
| QY | 185 | cagctgttacaagatatcgaaagccttgcgaactggaagttagtctgtaacaactatcc | 244 | | |
| Db | 638 | CAAAATTCACATGACATTGAAAGTTTACGTAATTTAAAGTTCAACCATTAATCT | 697 | | |
| QY | 245 | aacctctttaaanaaacagtcagtaacccttaacaacacactcaagaacagcatatctgag | 304 | | |
| Db | 698 | GATATTTCAGAAACATTCGTGAAACCTTTAAAGAAGAAAGATGATTCATCGTTA----- | 751 | | |
| QY | 305 | gattcaataatgcacacacagtgtaagcgttagttcattcgtgcgcagttcagcgtcattacg | 364 | | |
| Db | 752 | --ATGATTTACACCAACCAATTCAGTGTAGTACGACACCAACAAAGAAAGAACATGCCAAG | 808 | | |
| QY | 365 | ccagtaactcttcaaccagaagtagttatttcaaccagtcgtagctgcctactcc | 424 | | |
| Db | 809 | ATTGTCATTCCTAAAGGTGCTTGTGAAACAAATGCTACATCTCTAGAAACCTCTATATCG | 868 | | |
| QY | 425 | ctgaaattacagacagcagattgtcataaagtggttagagattgcttccttaccagtaacaa | 484 | | |
| Db | 869 | AATAGCTTTACTGAAACAAATTAATTCGCGATGCGAATGGAACAAAGGTAATTCATTAATGAG | 928 | | |
| QY | 485 | cgagacatlaatgtgaacttggaatttaataagaacatlaagaacttaacttccaaat | 544 | | |
| Db | 929 | AGATGATCATCATGTTAATTTAGCAATTCATTAATTAACATCTTGAGACATGTTGATCCAACT | 988 | | |
| QY | 545 | tttgaagctgcctcatgtctgtaggttattatctgcgtcagtaatgataattga | 600 | | |
| Db | 989 | TTTGAAACTGTGTATGTTGTGATTTTATTTATTTGTTAGTGAACATTTAAATTTCA | 1044 | | |
| RESULT | 4 | | | | |
| CNS06NM5/c | CNS06NM5 | 1134 bp | DNA | GSS | 04-JUL-2001 |
| LOCUS | T7 end of clone AU00AA015C11 of library AU00A from strain CBS 3082 | | | | |
| DEFINITION | Of Saccharomyces kluyveri, genomic survey sequence. | | | | |
| ACCESSION | AL040694.7 | | | | |
| VERSION | AL040694.7.1 GI:12171645 | | | | |
| KEYWORDS | GSS. | | | | |
| SOURCE | Saccharomyces kluyveri. | | | | |
| ORGANISM | Saccharomyces kluyveri | | | | |
| REFERENCE | Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; | | | | |
| AUTHORS | 1 (bases 1 to 1134) | | | | |
| TITLE | Neuveglise, C., Bon, E., Lepingle, A., Winker, P., Attiguenave, F., | | | | |
| JOURNAL | Galliaridin, C., and Casaregola, S. | | | | |
| MEDLINE | Genomic exploration of the hemiascomycetous yeasts: 9. | | | | |
| REFERENCE | Saccharomyces kluyveri | | | | |
| AUTHORS | FEBS Lett. 487 (1), 56-60 (2000) | | | | |
| | 20584719 | | | | |
| | 2 (bases 1 to 1134) | | | | |
| | Souciet, U. L., Aigle, M., Attiguenave, F., Blandin, G., | | | | |
| | Biolotin-Fukuhara, M., Bon, E., Brothier, P., Casaregola, S., | | | | |
| | de-Montigny, J., Dujon, B., Durieux, P., Lepingle, A., Llorente, B., | | | | |
| | Majpetty, A., Neuveglise, C., Ozier-Kalogeropoulos, O., Potier, S., | | | | |
| | Saurin, W., Tekata, F., Toffano-Niccho, C., Wesolowski-Louvel, M., | | | | |
| | Winker, P., and Weissenbach, J. | | | | |
| | Genomic exploration of the hemiascomycetous yeasts: 1. A set of | | | | |
| | yeast species for molecular evolution studies | | | | |
| | FEBS Lett. 487 (1), 3-12 (2000) | | | | |

| | | | |
|--------------|--|---|-------------|
| Db | 82 | AATACGTTTTCATTTTTTCCAGACTGCATACACTCAATAATTTGGTGTTTCACAA | 23 |
| Oy | 1693 | gttttattcgtgaagaacagg | 1713 |
| | | | |
| Db | 22 | CTTTATCTTCATCAGAACAGG | 2 |
| RESULT | 6 | | |
| CNS06PBO/c | | | |
| LOCUS | CNS06PBO | 933 bp | DNA |
| DEFINITION | T7 end of clone AV0AA012H1 of library AV0AA from strain CBS 379 of | GSS | 05-JUL-2001 |
| ACCESSION | Saccharomyces exiguus, genomic survey sequence. | | |
| VERSION | AL409138 | | |
| KEYWORDS | GI:12176372 | | |
| SOURCE | GSS. | | |
| ORGANISM | Saccharomyces exiguus. | | |
| REFERENCE | Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; | | |
| AUTHORS | Saccharomycetales; Saccharomycetaceae; Saccharomyces. | | |
| | Bon,E., Neuveglise,C., Lepingle,A., Wincker,P., Attiguenave,F., | | |
| | (bases 1 to 933) | | |
| TITLE | Gallardin,C. and Casaregola,S. | | |
| JOURNAL | Genomic exploration of the hemiascomycetous yeasts : 6. | | |
| MEDLINE | Saccharomyces exiguus | | |
| REFERENCE | FEBS Lett. 487 (1), 42-46 (2000) | | |
| AUTHORS | 20584716 | | |
| | 2 (bases 1 to 933) | | |
| | Souciet,J.L., Aigle,M., Attiguenave,F., Blandin,G., | | |
| | Bojotin-Fukuhara,M., Bon,E., Brotlier,P., Casaregola,S., | | |
| | de-Montigny,J., Dujon,B., Durrens,P., Lepingle,A., Lorente,B., | | |
| | Maupertuy,A., Neuveglise,C., Ozler-Kalogeropoulos,O., Potier,S., | | |
| | Saurin,M., Tekala,F., Toffano-Nioche,C., Weslowski-Louvel,M., | | |
| | Wincker,P. and Weissenbach,J. | | |
| TITLE | Genomic exploration of the hemiascomycetous yeasts : 1. A set of | | |
| JOURNAL | yeast species for molecular evolution studies | | |
| MEDLINE | FEBS Lett. 487 (1), 3-12 (2000) | | |
| REFERENCE | 20584711 | | |
| AUTHORS | 3 (bases 1 to 933) | | |
| TITLE | Genoscope. | | |
| JOURNAL | Direct Submission | | |
| | Submitted (07-SEP-2000) Genoscope - Centre National de Sequencage, | | |
| | 2 rue Gaston Cremlieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail : | | |
| | seqref@genoscope.cns.fr Web : www.genoscope.cns.fr) | | |
| COMMENT | This GSS is part of a random genomic sequencing program of thirteen | | |
| | yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces | | |
| | exiguus, Saccharomyces servazzii, Zygosaccaromyces rouxii, | | |
| | Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces | | |
| | lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia | | |
| | angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila, | | |
| | Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to | | |
| | 5 kb were prepared and both extremities were sequenced. See | | |
| | keywords for description of this sequence and for the sequence of | | |
| | the other extremity of this insert. | | |
| FEATURES | Location/Qualifiers | | |
| SOURCE | 1..933 | | |
| | /organism="Saccharomyces exiguus" | | |
| | /strain="CBS 379" | | |
| | /db_xref="taxon:34358" | | |
| | /clone="AV0AA012H11" | | |
| | /clone_lib="AV0AA" | | |
| | /note="end : 17" | | |
| | complement(1..11..>856) | | |
| | /note="similar to Saccharomyces cerevisiae ORF YML11w [| | |
| | strong similarity to ubiquitination protein Bulp j]" | | |
| | /evidence-not_experimental | | |
| misc_feature | complement(137..>932) | | |
| | /note="similar to Saccharomyces cerevisiae ORF YMR275c [| | |
| | Bul1 : ubiquitination pathway protein] | | |
| | 1 putative frameshift(s)" | | |
| | /evidence-not_experimental | | |
| BASE COUNT | 312 a 141 c 130 g 347 t 3 others | | |
| ORIGIN | | | |

| Query Match | 2.9% | Score 85.6 | DB 13 | Length 933 |
|-----------------------|--|--|----------|------------|
| Best Local Similarity | 57.5% | Pred. No. 4e-05 | | |
| Matches 154 | Conservative 0 | Mismatches 114 | Indels 0 | Gaps 0 |
| OY | 443 | atgttcataaattggttagagatgtgctcccttacagtaacagagacatlaattgtgaac | 502 | |
| Db | 346 | ATTGTAAGAGTGTGOTCACAAGTTGGACACACAGAATTTGAAAGATTTGCTTCGCAGAT | 287 | |
| OY | 503 | ttggaattaaataagagacattaaggaaccttaattccaaattttgaaagctgcctatgt | 562 | |
| Db | 286 | TTAGAAAGACTCGTGATATTGGAGAAACCTTAATTCACAACTTTGAAAGTTGCTTAGCT | 227 | |
| OY | 563 | tgtgaagtttcttggttcgagtaatgataatttgaataaccctcttggtgagcgaag | 622 | |
| Db | 226 | TGTGCTTTCTATGCACATTAGAGTTATATTTAAATTTGCTATGTCACATCGGATCAGCCAGT | 167 | |
| OY | 673 | atgatalcccatcttctcgttagagcaagtgcacaaataaaaaaacattagaaaaaatct | 682 | |
| Db | 166 | CTTGATATTCACAAATCAATGTTAAAGATGATATTTTGTCGATACGATTAAGCAAGCAT | 107 | |
| OY | 683 | cgtaactcttcttctagatagata 710 | | |
| Db | 106 | AGATACATATTTTATTTTATTTTAAATTTTA 79 | | |
| RESULT 7 | | | | |
| LOCUS | CNS061MU | 1111 bp | DNA | GSS |
| DEFINITION | T3 end of clone XAT0A002B08 of library XAT0A from strain CBS 4311 | | | |
| ACCESSION | AL404740 | | | |
| VERSION | AL404740.1 GI:12166586 | | | |
| KEYWORDS | GSS. | | | |
| SOURCE | Saccharomyces servazzii. | | | |
| ORGANISM | Saccharomyces servazzii. Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; Saccharomycetales; Saccharomycetaceae; Saccharomyces. | | | |
| REFERENCE | 1 (bases 1 to 1111) Casaregola,S., Lepingle,A., Bon,E., Neuveglise,C., Nguyen,H., Artiguenave,F., Wincker,P. and Gaillardin,G. Genomic exploration of the hemiascomycetous yeasts: 7. Saccharomyces servazzii FEBS Lett. 487 (1), 47-51 (2000) | | | |
| AUTHORS | 2 (bases 1 to 1111) Souciet,J.L., Aigle,M., Artiguenave,F., Blandin,G., Bolteln-Pukunara,M., Bon,E., Brothier,P., Casaregola,S., de-Montigny,J., Dujon,B., Durrens,P., Lepingle,A., Florente,B., Malpettuy,A., Neuveglise,C., Ozler-Kalopoulou,O., Potier,S., Saurin,W., Tekala,F., Toffano-Nioche,C., Wesolowski-Louvel,M., Wincker,P. and Weissenbach,J. Genomic exploration of the hemiascomycetous yeasts: 1. A set of yeast species for molecular evolution studies FEBS Lett. 487 (1), 3-12 (2000) | | | |
| JOURNAL | 20584717 | | | |
| MEDLINE | | | | |
| REFERENCE | | | | |
| AUTHORS | 3 (bases 1 to 1111) Genoscope. Direct Submission Submitted (07-SEP-2000) Genoscope - Centre National de Sequencage, 2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail : seque@genoscope.cns.fr - Web : www.genoscope.cns.fr) This GSS is part of a random genomic sequencing program of thirteen yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii, Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila, Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to 5 kb were prepared and both extremities were sequenced. See keywords for description of this sequence and for the sequence of the extremity of this insert. Location/Qualifiers | | | |
| JOURNAL | | | | |
| MEDLINE | | | | |
| REFERENCE | | | | |
| AUTHORS | | | | |
| TITLE | | | | |
| COMMENT | | | | |

```

source
  1. 1111
  /organism="Saccharomyces servazzii"
  /strain="CBS 4311"
  /db_xref="taxon:27293"
  /clone_lib="XAT0A02B08"
  /clone_id="XAT0A"
  /note="end : T3"
  <2..>609

misc_feature
  /note="similar to Saccharomyces cerevisiae ORF YML110c [
  DB156 : ubiquinone biosynthesis, methyltransferase ]
  1 putative frameshift(s)"
  /evidence="not_experimental"
  complement(<771..>938)
  /note="similar to Saccharomyces cerevisiae ORF YML111w [
  strong similarity to ubiquitination protein Bulp ]"
  /evidence="not_experimental"

misc_feature
  /note="similar to Saccharomyces cerevisiae ORF YML111w [
  strong similarity to ubiquitination protein Bulp ]"
  /evidence="not_experimental"

BASE COUNT      376 a      157 c      181 g      382 t      15 others
ORIGIN

Query Match      2.8%: Score 83; DB 13; Length 1111;
Best Local Similarity 62.1%: Pred. No. 0.0001;
Matches 131; Conservative 0; Mismatches 80; Indels 0; Gaps 0;

QY 492 ttaatgtaacttggaacttaataaagacataaagaaacttaattccagtttgaa 551
    ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 937 TTAATGTTAACTTCATGCTTAATCAAAATATATAATGTTACATATATTCCTATTGAAA 878

QY 552 gctgcatactgttaggtttattatgcgttcgaataagattaaattgaaacacttg 611
    || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 877 CATGTTTATGTTGATGATTTATTTCTTCGTTAAATGTCATAATTCGATATATATATTG 818

QY 612 gcgttagcgaagatgatataccctattctgttagcgaagtgcgaataaataaacaacta 671
    || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 817 GTTCCACAGACATGCGATATCCAGTTAATGTTAAGAACTCATATATAGTGCGAAGAAAGG 758

QY 672 gaaaaaattcgttactttctataagata 702
    || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 757 CATAGCACTTTTATAGAAATTTAGTAAGTA 727

RESULT 8
CNS0039G      1101 bp      DNA      GSS      03-JUN-1999
LOCUS
DEFINITION   Drosophila melanogaster genome survey sequence TET3 end of BAC #
              BACR08K10 of RPCI-98 library from Drosophila melanogaster (fruit
              fly) genomic survey sequence.
ACCESSION   AL063921
VERSION     AL063921.1 GI:4941778
KEYWORDS    GSS.
SOURCE      fruit fly.
ORGANISM    Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachyoptera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1 (bases 1 to 1101)
AUTHORS     Genoscope.
TITLE       Direct Submission
JOURNAL     Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 Evry cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - web : www.genoscope.cns.fr)
            Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mamoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain Y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library

```

```

and how to order individual BAC clones, the entire library, or
filters for hybridization from the BACAC Resource Center can be
found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES
  source
    1. 1101
    /organism="Drosophila melanogaster"
    /db_xref="taxon:7227"
    /clone_lib="RPCI-98"
    /clone_id="BACR08K10"
    /note="end : TET3"

BASE COUNT      201 a      64 c      131 g      202 t      503 others
ORIGIN

Query Match      2.7%: Score 79.4; DB 13; Length 1101;
Best Local Similarity 19.7%: Pred. No. 0.00042;
Matches 139; Conservative 287; Mismatches 273; Indels 6; Gaps 2;

QY 611 ggcgtagcgaagatgatataccctattctgttagcgaagtgcgaataaataaacaact 670
    || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 386 GGRGGGGGGMATATAMAMMMTTTTTTTAAAMAAAATATATTTAAAMAAAATTT 445

QY 671 agaaaaattcgttactttctataagataagataatgtagtgcctatagat 730
    : | | | | : | | | | | | | | | | | | | | | | | | | | | | |
Db 446 WMAAAAMAAAMAMWTATMTATMAAAMAAAATTTTTTATTTATTTATTTATAT 505

QY 731 gaagattatcgcgcctttgattatccctattatataataaattccttaaatg--- 787
    || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 506 WTTMTTAAAMAAAAMAAAAMAAAAMAAAAMAAAATTTTATTTTAAAMATAAM 565

QY 788 -catttcgtgcctctctgttctgtctgtatttttttttttttttttttttttttttttt 846
    : | | | | : | | | | | | | | | | | | | | | | | | | | | | |
Db 566 CMAAMTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 625

QY 847 aaaccttgatgatttataccttattttaaatttaataataatgcagaatgcaga 906
    : | | | | : | | | | | | | | | | | | | | | | | | | | | | |
Db 626 MAMMMMTTATTAAYUUYUUYUUYUUYUUYUUYUUYUUYUUYUUYUUYUUYUUYUUYU 685

QY 907 caaaacataagatttcttcttgcgaagaaataaagaataatgtagtgcctgagta 966
    : | | | | : | | | | | | | | | | | | | | | | | | | | | | |
Db 686 MTCAMCMTCTGCHCCTUUYUUYUUYUUYUUYUUYUUYUUYUUYUUYUUYUUYU 745

QY 967 ctgactgtctgtagagaagaagaagaagaatgcgaataatgcgaataatgcgaat 1026
    : : : | | | | : | | | | | | | | | | | | | | | | | | | | | |
Db 746 NNNNNMAUUTUUYUUYUUYUUYUUYUUYUUYUUYUUYUUYUUYUUYUUYUUYUUY 805

QY 1027 gctctataacttccatatacctatgctgctacgctataatgcgaataatgcga 1086
    | : : | | | | | : | | | | | | | | | | | | | | | | | | | |
Db 806 TCCUUYUUYUUYUUYUUYUUYUUYUUYUUYUUYUUYUUYUUYUUYUUYUUYU 865

QY 1087 gctcaacagatatactcgcgaagataatgcgaataatgcgaataatgcgaatgct 1146
    | : : | | | | | : | | | | | | | | | | | | | | | | | | | |
Db 866 NATTTATATGTCACATTTTNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 923

QY 1147 aatttcaacatttttgcgtcttcttcaaatcaaatcaaatcaaatcaaatcaaat 1206
    : | | : | | | | | | | | | | | | | | | | | | | | | | | |
Db 924 NNTGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 983

QY 1207 tataagttcaagacgacttttaagcttgtagatatacccaagttatacgtgcact 1266
    | : : | | | | | : | | | | | | | | | | | | | | | | | | | |
Db 984 TATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 1043

QY 1267 gccaacaccttgcaatgcgaagctgcgtggaagacttaagca 1311
    : | | | | | : | | | | | | | | | | | | | | | | | | | |
Db 1044 ATAMACSTGNNNTGNTCTGTTTNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1088

RESULT 9
AL547503      1125 bp      mRNA      EST      16-FEB-2001
LOCUS
DEFINITION   AL547503 LTI_NF01006_P12 Homo sapiens cDNA clone CS001011YB03 5
              prime, mRNA sequence.
ACCESSION   AL547503

```


| | | |
|------------|--|-----------------------------|
| VERSION | AL547503.1 | GI:12881624 |
| KEYWORDS | EST. | |
| SOURCE | human. | |
| ORGANISM | Homo sapiens | |
| REFERENCE | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. | |
| AUTHORS | 1 (bases 1 to 1125) | |
| TITLE | Li, W.B., Gruber, C., Jesse, J. and Polayes, D. | |
| JOURNAL | Full-length cDNA libraries and normalization | |
| COMMENT | Unpublished (2001) | |
| FEATURES | Contact: Genoscope Genoscope - Centre National de Sequencage BP 191 91006 Evry cedex - France Email: seque@genoscope.cns.fr, Web : www.genoscope.cns.fr. | |
| SOURCE | location/Qualifiers | |
| | 1..1125 | |
| | /organism="Homo sapiens" | |
| | /db_xref="taxon:9606" | |
| | /clone="CSOD1011YB03" | |
| | /clone_1lb="LFL_NFL006_PL2" | |
| | /tissue_type="Placenta" | |
| | /note="Vector: pCMVSPORT 6; Site 1: NotI; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies. Contact : Peng Liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive, Rockville, Maryland 20850, USA Fax : (1) 301 610 8371 | |
| | Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com" | |
| BASE COUNT | 415 a | 142 c 148 g 345 t 75 others |
| ORIGIN | | |

| Query Match | 2.6% | Score 75.6 | DB 10 | Length 1125 |
|-----------------------|---|------------------|----------|-------------|
| Best Local Similarity | 43.9% | Pred. No. 0.0018 | | |
| Matches 138 | Conservative 32 | Mismatches 144 | Indels 0 | Gaps |
| OY 648 | aagtgacaaaataaaaaaacattgaagaaaattccgtacttctctatagatatagat | 707 | | |
| Db 759 | AAAAAAAAAAAAAAAAAATANNAAAAAATAANTNNAAAAATTTTWTATWAT | 818 | | |
| OY 708 | atattgacggcttcgttatagatgaaggattatcgcgctcttgatccctattt | 767 | | |
| Db 819 | NTWTWTATATTAATTTTWTWTWTWTWTWTWTWTWTWTWTWTWTWTWTWTWT | 878 | | |
| OY 768 | aataaatctctttaaataatgcattcttcggtcctcttcttgcttcgatactttt | 827 | | |
| Db 879 | TATATNAAAAATTTTATTTTWTWTWTWTWTWTWTWTWTWTWTWTWTWTWTWT | 938 | | |
| OY 828 | tttggaccactgtagtgaaaaccttgatgattatcattacatttaagtactaa | 887 | | |
| Db 939 | TTATATNAAAAAAMAAAAATATTTTWTWTWTWTWTWTWTWTWTWTWTWTWT | 998 | | |
| OY 888 | aatacgcgaatttcagsgaacaacacatgaattctcttgcagaagaaaataacgaa | 947 | | |
| Db 999 | AAAAATATWATATTTTWTWTWTWTWTWTWTWTWTWTWTWTWTWTWTWTWT | 1058 | | |
| OY 948 | taaatctgactctt 961 | | | |
| Db 1059 | AAAAATTTTWTWTWTWTWTWTWTWTWTWTWTWTWTWTWTWTWT 1072 | | | |

RESULT 10

CNS00EVL/C

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

GSS.

GI:4949849

CNS00EVL 1101 bp DNA

Drosophila melanogaster genome survey sequence T7 end of BAC: BACR2B23 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.

AL069706

AL069706.1

GSS.

04-JUN-1999

| FEATURES | REFERENCE | COMMENT |
|----------|--|---------|
| ORGANISM | fruit fly. Drosophila melanogaster Euarysta, Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila. | |
| AUTHORS | 1 (bases 1 to 1101) Genoscope. | |
| TITLE | Direct Submission | |
| JOURNAL | Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : segre@genoscope.cns.fr - Web : www.genoscope.cns.fr) | |
| | Determination of this BAC-end and sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see http://www.fruitfly.org/TheBDGP/Drosophila_melanogaster_BAC_library was prepared by Kazutoyo Osoegawa and Aaron Mammossier at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCT-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm . | |
| | Location/Qualifiers | |

| BASE COUNT | 419 a | 91 c | 60 g | 299 t | 232 others |
|-----------------------|------------------|---|-------------------|-------|--------------|
| ORIGIN | | | | | |
| Query Match | | 2.6% | Score 75.2 | DB 13 | Length 1101; |
| Best Local Similarity | | 37.3% | Pred. No. 0.0021; | | |
| Matches 152; | Conservative 64; | Mismatches 192; | Indels 0; | Gaps | |
| Qy | 563 | tgtaggtttatcgcgltcgaagtaagatlaaalttgaanaaccatcttgcgtacggaag | 6222 | | |
| Db | 1050 | twfammwtatwtatattatatttwtataaataatatawtataatataatataatataat | 991d | | |
| Qy | 623 | atgataaccatattctgttagcgcaagtcgaacaaataaaaaacatctagaaaaaatct | 6682 | | |
| Db | 990 | ataataaanaatataamwtamaaamwtatamwtataaamaataamataamwtata | 9311 | | |
| Qy | 683 | cgtactcttctctagatagatagatagatgattgcttgcctctagatggaagattat | 7424 | | |
| Db | 930 | tt | 871j | | |
| Qy | 743 | cgcgctccttgatcccatatlaataaaatcttlaaaatgcatcttctgcgtc | 802d | | |
| Db | 870 | atwtatwtatattatatt | 8111 | | |
| Qy | 803 | tttctgtcctctgtaatt | 862z | | |
| Db | 810 | atttatwtwtwtwtatatt | 751 | | |
| Qy | 863 | tattaccttattttaaagttactaaataatcagagattcaggaacaaacatgaaattc | 9222 | | |
| Db | 750 | wmwtatwtwtwtwtatatt | 691d | | |
| Qy | 923 | cttgcgtcaagaanaataaagaataaatatgactgcttgcactctga | 970 | | |
| Db | 690 | wtatttttaaaattatttttaattatamwaaawmwtatttttaataamwtatatta | 643 | | |
| RESULT 11 | | | | | |
| CNS00FUH/c | | | | | |
| LOCUS | CNS00FUH | 996 bp | DNA | GSS | 03-JUN-1999 |

DEFINITION Drosophila melanogaster genome survey sequence TET3 end of BAC: BACR31021 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.

ACCESSION AL071063

VERSION AL071063.1 GI:4951105

KEYWORDS GSS.

SOURCE fruit fly.

ORGANISM Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

REFERENCE 1 (bases 1 to 996)

AUTHORS Genoscope.

TITLE Direct Submission

JOURNAL Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)

COMMENT Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see <http://www.fruitfly.org> The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammoler in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's pl and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

location/Qualifiers

source 1. 996

/organism="Drosophila melanogaster"

/db_xref="taxon:7227"

/clone_lib="RPCI-98"

/clone="BACR31021"

/note="end : TET3"

BASE COUNT 383 a 164 c 81 g 171 t 197 others

ORIGIN

Query Match 2.4%; Score 70.8; DB 13; Length 996;

Best Local Similarity 37.4%; Pred. No. 0.012;

Matches 151; Conservative 71; Mismatches 176; Indels 6; Gaps 1;

QY 620 aagattgataccctatttcgttagcgaagtagcaaaataaataacattagaataat 679

DB 945 WAAATTTWTTAATTTBTBTBGGSSGSSGSSGSGRRRAAAAAAAMWMAAT 886

QY 680 tctgtaaccttctatagatagatatagtggtcgtcatagatgaagtgat 739

DB 885 TTTTWTAAATTTAARTTAAATTTAAATTTTATTTTATTTTATTTTATTTT 826

QY 740 taccgctccttgcataccatataataaattcttaaatgcatcttcgtg 799

DB 825 WTTTATTTTAAATTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTT 772

QY 800 cttcttgcctcctgcatcttcttcttcttgcgaacagcgagaaaccttgatga 859

DB 771 TTTTATTTTAAATTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTT 712

QY 860 tttatataccttcttaagttactaaataatgagatcaggaacaaacatagat 919

DB 711 TTTTATTTTAAATTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTT 652

QY 920 tttcttgcagaataaataaataaataatgagatccttgcactgagctgctc 979

DB 651 TTTTATTTTAAATTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTT 592

QY 980 atagagagaccagagacagcaatgctacaatcttcaacattctc 1023

Db 591 TMAAATATKADTWAMTTTAAATTTTATTTATTTATTTATTTT 548

RESULT 12

CNS001FB 1101 bp DNA GSS 03-JUN-1999

LOCUS

DEFINITION Drosophila melanogaster genome survey sequence TET3 end of BAC # BACR04A23 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.

ACCESSION AL060732

VERSION AL060732.1 GI:4939397

KEYWORDS GSS.

SOURCE fruit fly.

ORGANISM Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

REFERENCE 1 (bases 1 to 1101)

AUTHORS Genoscope.

TITLE Direct Submission

JOURNAL Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)

COMMENT Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see <http://www.fruitfly.org> The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammoler in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's pl and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

location/Qualifiers

source 1. 1101

/organism="Drosophila melanogaster"

/db_xref="taxon:7227"

/clone_lib="RPCI-98"

/clone="BACR04A23"

/note="end : TET3"

BASE COUNT 288 a 110 c 103 g 491 t 109 others

ORIGIN

Query Match 2.3%; Score 66.8; DB 13; Length 1101;

Best Local Similarity 38.8%; Pred. No. 0.056;

Matches 165; Conservative 55; Mismatches 199; Indels 6; Gaps 1;

QY 532 ttaattccaagtttgaagcctgctatggttaggttattatgctgctgagtaagt 591

DB 675 TTTTATTTTAAATTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTT 734

QY 592 taatttgaacacatcttgcgtagaagaatgatacccatcttcttgcgaagt 651

DB 735 WAAATTTTAAATTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTT 794

QY 652 gacaaataaataaataaataaataatcctgcttcttctatagatagatat 711

DB 795 ATTAATTTAAATTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTT 848

QY 712 gtagtgcttctatagatgaaggtattatcgcgtccttctgattccattat 771

DB 849 ATAAATTTAAATTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTT 908

QY 772 aaattctttaaataatgcatcttctgctgcttcttctgcttcttcttctt 831

DB 909 AATTTAAATTTAAATTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTT 968

COMMENT

Determination of this BAC-end sequence was carried out as part of collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see <http://www.fruitfly.org> The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Oosagawa and Aaron Mammossier in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

| | | | | |
|-----------------------|-----------------|-----------------|----------|-------------|
| Query Match | 2.3% | Score 66.8 | DB 13 | Length 1101 |
| Best Local Similarity | 36.5% | Pred. No. 0.056 | | |
| Matches 216 | Conservative 70 | Mismatches 301 | Indels 5 | Gaps 1 |

| | | | |
|----|------|---|------|
| DQ | 650 | gtagacaaaataaaaaaacattagaaaaaatctcgtccatcttcctatagaataagatc | 709 |
| | | : : : : : : : : | |
| Dd | 1091 | RMGNGCATGAMGTWMDKFWMTTWTATNADTTKTCTTTTTTKMTATRMATAKAAKMTTT | 1032 |
| QY | 710 | atgatggttgcgtcatagaaagagcatltaacgcgcctcltgyatccactaatcaa | 769 |
| | | : : : : : : : | |
| Dd | 1031 | TWGFRTWTTRTWMTWMNATNTATTATTTTTTTTTTMMWTATTAATCAATVATATA | 972 |
| QY | 770 | taaaatctctttaaatacgatcttctcgtgc-----tctttgcgtgcctcgtatatttt | 824 |
| | | : : : : : : : : : : | |

| | |
|------------|--|
| RESULT | 14 |
| AL514935 | |
| LOCUS | 619 bp mRNA EST 13-FEB-2001 |
| DEFINITION | AL514935 LTI NPL006.PL2 Homo sapiens cDNA clone Cl08B012XD08 3 prime. mRNA sequence. |
| ACCESSION | AL514935 |
| VERSION | AL514935.1 GI:12778428 |
| KEYWORDS | EST. |
| SOURCE | human. |
| ORGANISM | Homo sapiens |
| | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. 1 (bases 1 to 619) Li,W.B., Gruber,C., Jesse,J. and Polayes,D. Full-length cDNA libraries and normalization unpublished (2001) |
| REFERENCE | Contact: Genoscope |
| AUTHORS | Genoscope, Centre National de Sequencage |
| JOURNAL | Bp 191 91006 EVRI cedex - France |
| COMMENT | Email: segr@genoscope.cns.fr, Web : www.genoscope.cns.fr. |

was primed with a NotI-colligo(dT) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies Contact Feng Liang Life Technologies, a division of Invitrogen 9600 Medical Center Drive, Rockville, Maryland 20850, USA Fax: (1) 301 610 8371 Email: liang@lifetech.com URL: <http://fulllength.invitrogen.com>

| | | | | | |
|-------------|------|------|-------------|--------|-------------|
| BASE COUNT | 52 A | 8 C | 10 G | 314 T | 33 OTHERS |
| ORIGIN | | | | | |
| Query Match | | 2.3% | Score 66.6; | DB 10; | Length 619; |

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 12, 2002, 15:33:56 : Search time 91.73 Seconds
(without alignments)
7278.504 Million cell updates/sec

Title: US-09-165-460a-3

Perfect score: 2948

Sequence: 1 tgaactgttgatgaacaag.....gggagagataaagaatcacac 2948

Scoring table: IDENTITY_NUC

Gapop 10.0, Gapext 1.0

Searched: 351203 seqs, 11323899 residues

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A.COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B.COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A.COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B.COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCRTUS.COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|--------------------|
| C 1 | 48.2 | 1.6 | 6243 | 2 | US-09-056-075-1 |
| C 2 | 48 | 1.6 | 19124 | 2 | US-08-487-826B-13 |
| C 3 | 47.8 | 1.6 | 1117 | 4 | US-09-247-373B-33 |
| C 4 | 47 | 1.6 | 2447 | 2 | US-09-014-969-14 |
| C 5 | 46.4 | 1.6 | 731 | 1 | US-08-451-405A-2 |
| C 6 | 46 | 1.6 | 240 | 1 | US-08-628-417-6 |
| C 7 | 45 | 1.5 | 3012 | 2 | US-08-475-427-5 |
| C 8 | 45 | 1.5 | 3012 | 2 | US-07-842-165-5 |
| C 9 | 44.6 | 1.5 | 636 | 4 | US-08-998-416-1137 |
| C 10 | 44.6 | 1.5 | 1483 | 4 | US-09-262-749-1 |
| C 11 | 44.2 | 1.5 | 5852 | 1 | US-07-867-106-2 |
| C 12 | 43.6 | 1.5 | 1230 | 3 | US-09-022-669-3 |
| C 13 | 43.6 | 1.5 | 19124 | 2 | US-08-487-826B-13 |
| C 14 | 43 | 1.5 | 1553 | 3 | US-09-022-669-1 |
| C 15 | 42.8 | 1.5 | 7218 | 1 | US-08-232-463-14 |
| C 16 | 42.6 | 1.4 | 837 | 4 | US-08-998-416-288 |
| C 17 | 42.6 | 1.4 | 3138 | 1 | US-07-867-106-4 |
| C 18 | 42.4 | 1.4 | 578 | 4 | US-09-020-956-104 |
| C 19 | 42.4 | 1.4 | 578 | 4 | US-09-030-607-104 |
| C 20 | 42 | 1.4 | 2110 | 4 | US-09-419-459-1 |
| C 21 | 42 | 1.4 | 5852 | 1 | US-07-867-106-2 |
| C 22 | 41.6 | 1.4 | 4376 | 1 | US-08-119-125A-1 |
| C 23 | 41.6 | 1.4 | 6744 | 1 | US-08-119-125A-2 |
| C 24 | 41.2 | 1.4 | 1474 | 4 | US-08-821-994-64 |
| C 25 | 40.8 | 1.4 | 2230 | 3 | US-08-844-188-34 |
| C 26 | 40.4 | 1.4 | 8920 | 2 | US-08-446-855A-1 |
| C 27 | 40.4 | 1.4 | 8920 | 4 | US-09-150-741-1 |

| | | | | | | |
|------|------|-----|------|---|-------------------|--------------------|
| C 28 | 40.2 | 1.4 | 658 | 4 | US-08-998-416-595 | Sequence 595, App |
| C 29 | 40.2 | 1.4 | 686 | 4 | US-08-953-326-21 | Sequence 21, Appl |
| C 30 | 40.2 | 1.4 | 1641 | 1 | US-08-300-903A-8 | Sequence 8, Appl |
| C 31 | 40.2 | 1.4 | 4970 | 1 | US-08-764-100-14 | Sequence 14, Appl |
| C 32 | 40.2 | 1.4 | 4970 | 1 | US-08-764-100-14 | Sequence 20, Appl |
| C 33 | 40 | 1.4 | 7218 | 1 | US-08-232-463-14 | Sequence 14, Appl |
| C 34 | 39.8 | 1.4 | 688 | 6 | 5498694-3 | Patent No. 5498694 |
| C 35 | 39.8 | 1.4 | 960 | 3 | US-09-248-335-57 | Sequence 57, Appl |
| C 36 | 39.8 | 1.4 | 1602 | 1 | US-08-530-950-3 | Sequence 3, Appl |
| C 37 | 39.8 | 1.4 | 1602 | 1 | US-08-888-429A-3 | Sequence 3, Appl |
| C 38 | 39.8 | 1.4 | 1602 | 4 | US-09-149-879-3 | Sequence 3, Appl |
| C 39 | 39.4 | 1.3 | 1243 | 2 | US-08-687-080-53 | Sequence 53, Appl |
| C 40 | 39.4 | 1.3 | 1582 | 3 | US-08-545-196B-10 | Sequence 10, Appl |
| C 41 | 39.4 | 1.3 | 1582 | 3 | US-08-545-196B-12 | Sequence 12, Appl |
| C 42 | 39.4 | 1.3 | 5893 | 1 | US-08-592-126-54 | Sequence 54, Appl |
| C 43 | 39.4 | 1.3 | 5893 | 2 | US-08-687-080-44 | Sequence 44, Appl |
| C 44 | 39.2 | 1.3 | 926 | 3 | US-08-945-994-8 | Sequence 8, Appl |
| C 45 | 39.2 | 1.3 | 2058 | 2 | US-08-749-391-1 | Sequence 1, Appl |

ALIGNMENTS

RESULT 1
US-09-056-075-1/c
Sequence 1, Application US/09056075
Patent No. 595368
GENERAL INFORMATION:
APPLICANT: Johnson, Eric A.
APPLICANT: Bradshaw, Marile
APPLICANT: Rood, Julian
TITLE OF INVENTION: Expression System for Clostridium
TITLE OF INVENTION: Species
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Quarles & Brady
STREET: 1 South Pinckney Street
City: Madison
STATE: WI
COUNTRY: US
ZIP: 53701-2113
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/056,075
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Seay, Nicholas J.
REGISTRATION NUMBER: 27386
REFERENCE/DOCKET NUMBER: 960296, 95238
TELECOMMUNICATION INFORMATION:
TELEPHONE: 608-251-5000
TELEFAX: 608-251-9166
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 6243 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: misc feature
LOCATION: 3770..4013
OTHER INFORMATION: /note= "RP4 origin of DNA transfer (orig) from
US-09-056-075-1."
Query Match 1.6%; Score 48.2; DB 2; Length 6243;

| | | | | |
|---------------------------|--------|------------------|-----------|---------------|
| Query Match | 1.68; | Score 48; | DB 2; | Length 19124; |
| Best Local Similarity | 51.9%; | Pred. NO. 0.032; | | |
| Matches 108; Conservative | 0; | Mismatches 100; | Indels 0; | Gaps 0; |

RESULT 4
US-09-014-969-14/c
: Sequence 14, Application US/09014969

```

: Patent No.5965397
: GENERAL INFORMATION:
: APPLICANT: Jacobs, Kenneth
: APPLICANT: McCoy, John M.
: APPLICANT: Lavalley, Edward R.
: APPLICANT: Racie, Lisa A.
: APPLICANT: Merberg, David
: APPLICANT: Treacy, Maurice
: APPLICANT: Spaulding, Vikki
: TITLE OF INVENTION: AGOSTILIN, Michael J.
: TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES
: NUMBER OF SEQUENCES: 32
: CORRESPONDENCE ADDRESSES:
: ADDRESSEE: Genetics Institute, Inc.
: STREET: 87 Cambridgepark Drive
: CITY: Cambridge
: STATE: MA
: COUNTRY: U.S.A.
: ZIP: 02140
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentln Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/014,969
: FILING DATE:
: CLASSIFICATION:
: ATTORNEY/AGENT INFORMATION:
: NAME: Sprunger, Suzanne A.
: REGISTRATION NUMBER: 41,323
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (617) 498-8284
: TELEFAX: (617) 876-5851
: INFORMATION FOR SEQ ID NO: 14:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 2447 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
: MOLECULE TYPE: cDNA
US-09-014-969-14

Query Match 1.6%; Score 47; DB 2; Length 2447;
Best Local Similarity 49.4%; Pred. No. 0.026;
Matches 119; Conservative 1; Mismatches 121; Indels 0; Gaps 0.

QY 679 ttctcgtaactttcctaagatagatatagatggttgcttatagatgaaggtaa
Db 2446 tttttttttttttttttttttttttttttttttttttttttttttttttttt
QY 739 ttatgcgccttcgtatcccatattataaaaaaatcctttaaaagcaatttcggc
Db 2386 tttttttttttttttttttttttttttttttttttttttttttttttttttt
QY 799 gctcttgcctgcctcgtatatttttttttttttttttttttttttttttttttt
Db 2326 tttttttttttttttttttttttttttttttttttttttttttttttttttt
QY 859 attattacaccttatttaagttaacataaacatcgaagattccaagaacaacatagaa
Db 2266 tttttttttttttttttttttttttttttttttttttttttttttttttttt
QY 919 t c 919
Db 2206 T 2206

RESULT 5
US-08-451-405A-2/c
: Sequence 2, Application US/08451405A
```

```

Patent No.5736358
GENERAL INFORMATION:
APPLICANT: FASEL, NICOLAS JOSEPH
APPLICANT: REYMOND, CHRISTOPHE DOMINIQUE
TITLE OF INVENTION: DICTYOSTELID EXPRESSION VECTOR AND
TITLE OF INVENTION: METHOD FOR EXPRESSING A DESIRED PROTEIN
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESSEE: THE WEBB LAW FIRM
STREET: 700 KOPPERS BUILDING, 436 SEVENTH AVENUE
CITY: PITTSBURGH
STATE: PENNSYLVANIA
COUNTRY: UNITED STATES OF AMERICA
ZIP: 15219-1818
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" FLOPPY DISK
COMPUTER: Midwest Micro 486-50
OPERATING SYSTEM: DOS
SOFTWARE: WORDPERFECT 6.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/451,405A
FILING DATE: 26-MAY-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/965,273
FILING DATE: 15-JAN-1993
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 731
TYPE: NUCLEIC ACID
STRANDNESS: SINGLE
TOPOLOGY: UNKNOWN
US-08-451-405A-2

Query Match      1.6%   Score 46.4; DB 1; Length 731;
Best Local Similarity 53.3%; Pred. No. 0.023;
Matches 98; Conservative 0; Mismatches 86; Indels 0; Gaps 0;

Db      Qy      Db      Qy      Db      Qy      Db      Qy      Db      Qy      Db      Qy      Db      Qy      Db      Qy
648 aagtcacaaataaaaaaacattgagaataaatctcgttactttcttaagatagatagat 707
    |||||||         ||||| ||||| ||||| ||||| ||||| ||||| |||||
488 AATTAAAAAAATTTAAAAAATTTATTTTAAAAAAATAATTAATTAATTAATTAATTAATTA 429
    |||||||         ||||| ||||| ||||| ||||| ||||| ||||| |||||
708 atatgatgttgtctatagatagaaggatattatcgccgtcccttgtaattcccatatt 767
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
428 TTTTAACCTTGTAAGATGAATGAGGTGTGTTAAAAGTGTGTTGGCAATTAAGAATA 369
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
768 ataataattctttaaaaatgcatttcctgtgcctctttgttgcctcttgcgtatttttt 827
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
368 TATAATATATATTTTAAAAAATTTCTTAAATACCTTTTTTTTTTGATTTTAATTTTTTTT 309
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy      828 tttt 831
Db      308 TTTT 305

RESULT        6
US-08-628-417-6/c
Sequence 6, Application US/08628417
Patent No. 5627054
GENERAL INFORMATION:
APPLICANT: GILLESPIE, DAVID
TITLE OF INVENTION: COMPETITOR PRIMER ASYMMETRIC
TITLE OF INVENTION: POLYMERASE CHAIN REACTION
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESSEE: U.S. ARMY CHEMICAL AND BIOLOGICAL
ADDRESS: DEFENSE COMMAND
STREET: OFFICE OF THE CHIEF COUNSEL (AMSCB-GC)
CITY: ABERDEEN PROVING GROUND
STATE: MARYLAND
COUNTRY: USA
ZIP: 21010-5423
```

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/628,417
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: BIFRONI, ULYSSES J
REGISTRATION NUMBER: 39,908
TELECOMMUNICATION INFORMATION:
TELEPHONE: 410-671-1158
TELEFAX: 410-671-2534
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 240 bases
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: oligodeoxynucleotide
HYPOTHETICAL: NO
ANTI-SENSE: YES
US-08-628-417-6

Query Match 1.6% Score 46; DB 1; Length 240;
Best Local Similarity 51.5% Pred. No. 0.019;
Matches 106; Conservative 0; Mismatches 100; Indels 0; Gaps 0;

Qy 685 ttactttctatagatagatataatgacgttctttagatgaagttattatcg 744
Db 237 TTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTT 178
Qy 745 cgtcccttgatcccatatttaataaattcttaaaatgacgttcttgcgtctc 804
Db 177 TTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTT 118
Qy 805 ttgttgctcgtatttttttttttggaccacgcgtgaaacatttgatctta 864
Db 117 TTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTT 58
Qy 865 ttaccattttaaagttacaaat 890
Db 57 TTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTT 32

RESULT 7
US-08-475-427-5/c
Sequence 5, Application 'US/08475427
Patent No. 5859340
GENERAL INFORMATION:
APPLICANT: DUBOIS, Michel
APPLICANT: GRISON, Rene
APPLICANT: LEGUAY, Jean-Jacques
APPLICANT: PIGNARD, Annie
APPLICANT: TOPPAN, Alain
TITLE OF INVENTION: Recombinant gene coding for a protein
TITLE OF INVENTION: having endochitinase activity or for a precursor thereof.
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 3000 K Street, N.W., Suite 500
CITY: Washington, D.C.
COUNTRY: USA
ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/475,427
FILING DATE: 07-JUN-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/842,165
FILING DATE: 01-MAY-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/FR91/00607
FILING DATE: 21-JUL-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: FR 90 09460
FILING DATE: 24-JUL-1990
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 16781/564/BDL
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202)672-5300
TELEFAX: (202)672-5399
TELEX: 904136
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 3012 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: Intron
LOCATION: 2384..2462
OTHER INFORMATION: /number= 1
FEATURE:
NAME/KEY: Intron
LOCATION: 2617..2697
OTHER INFORMATION: /number= 2
FEATURE:
NAME/KEY: CDS
LOCATION: join(1942..2383, 2463..2616, 2698..3007)
FEATURE:
NAME/KEY: sig_peptide
LOCATION: 1942..2166
FEATURE:
NAME/KEY: mat_peptide
LOCATION: join(2167..2383, 2463..2616, 2698..3007, 2698)
US-08-475-427-5

Query Match 1.5% Score 45; DB 2; Length 3012;
Best Local Similarity 45.4% Pred. No. 0.087;
Matches 162; Conservative 0; Mismatches 195; Indels 0; Gaps 0;

Qy 1355 gatttgatataatcatttataatccaaagctcttacttgaaagattttaccat 1414
Db 970 GAATTTGTTTAAATGAATTAATTAATTAATTAATTAATTAATTAATTAATTA 911
Qy 1415 gaattcctgaattttggaagtttgaagaatttataatttgaccacaataacaggagaata 1474
Db 910 TGAATTTATGATTTATGATTTATGATTTATGATTTATGATTTATGATTTATGATTT 851
Qy 1475 tttaacagctcaagcttttgaactgaacttaaaccttaacccgaatttcgaactaagc 1534
Db 850 TGACAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 791
Qy 1535 tataaagttatttggcaacacgcgttttttggacttgcgcagcgaacacatgct 1594
Db 790 GCCCAATATTATTGTTGTTGATGATGAACTTAATTTCTACGTGAATCAAAATATAGTGT 731
Qy 1595 tatgaagaatttaacaggaagcttcacatgaacacgtttccatcttcgcgcgaacatgcttc 1654
Db 730 AATGTGCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 671
Qy 1655 caatttatacacaacattttggagggttaaccaaatttgcgtatcgttaagaca 1711

ORIGINAL SOURCE:
ORGANISM: PAC1692RP
US-08-998-416-1137

Query Match 1.5%: Score 44.6; DB 4; Length 636;
Best Local Similarity 44.1%: Pred. No. 0.06;
Matches 230; Conservative 0; Mismatches 289; Indels 2; Gaps 1;

384 agagtagttatcttccctgagcgtagctgcttactccctgaattccagaccga 443
12 ATATAATTTATTTATTAAGATTATTAATTAACCTTTTATTTATTAATTAAGATTAA 71
444 tctgcaaaatgggttagagttgctcccttacagtcacaaagagacatcgtgaact 503
72 TTATTTAACATATTATATCATTTATTAATTAATTAATTAATTAATTAATTAATTT 131
504 tggaaatttaagaagacatgaagaacatcattccaaagtttgaagcctgctatgt 563
132 ATATAATTTATTTATTAATTTACTTAATTCATCATTTATTAATTAATTAATTA 191
564 gttagtttattgctgcgagtaagtaattgaataaccactggtgctagcgaaga 623
192 TAAATTTAATTAATGAATTAATTAATTAATTAATTAATTAATTAATTAATTA 251
624 tggatccctcattcctgtagcgaagtgacaaataaaacattgaa--aaatc 681
252 TTATTTAAGATTTATTTATTTCTTAATTAATTAATTAATTAATTAATTAATTA 311
682 tctgactcttctatagatagatagatagatagatagatagatagatagatagat 741
312 TTATTTAATTTATTTATTTAATTAATTAATTAATTAATTAATTAATTAATTT 371
742 tctgagctcttctgattccctatataataaattctttaaagatcttccggtc 801
372 AAATATTGTAATTTATTTATTTATTTAATTAATTAATTTATTAATTAATTTG 431
802 ctcttgctgctcctgattcttcttcttctgacacacgagtggaacacttgatg 861
432 ATATTATTATTTATTTATTTATTAAGATTTATTTAAATTAATTTATTTAACTT 491
862 ttattacccttatttaagtaactaaataacatcgagattca 902
492 TTATTAAATTTTATTAATTTATTAATTAATTAATTAATTAATTAATTTA 532

RESULT 10
US-09-262-749-1
Sequence 1, Application US/09262749
Patent No. 6261793

GENERAL INFORMATION:
APPLICANT: Whyte, David
APPLICANT: McGillik, Marlene
APPLICANT: Nunez-Oliva, Irma
APPLICANT: Hockenderry, Tish
APPLICANT: Pal, James
TITLE OF INVENTION: RAS CONVERTING ENDOPROTEASE (RCE) AND
TITLE OF INVENTION: METHODS
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: Schering-Plough Corporation
STREET: 2000 Galloping Hill Road K-6-1-1990
CITY: Kenilworth
STATE: New Jersey
COUNTRY: USA
ZIP: 07033

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: Power Macintosh
OPERATING SYSTEM: MACOS
SOFTWARE: Microsoft Word
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/262,749

FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Thampoe, Immac J
REGISTRATION NUMBER: 36,322
REFERENCE/DOCKET NUMBER: OC01005
TELECOMMUNICATION INFORMATION:
TELEPHONE: (908) 298-5061
TELEFAX: (908) 298-5388
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1483 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: 14..1000
US-09-262-749-1

Query Match 1.5%: Score 44.6; DB 4; Length 1483;
Best Local Similarity 53.4%: Pred. No. 0.083;
Matches 117; Conservative 0; Mismatches 99; Indels 3; Gaps 1;

1560 cgcctttcttggagcttcgacgacacacatctatgagcaattacagaagctcca 1619
615 CGCTCTTTTGGAGCTTCCCATTTTCACATATTATTAGACAGCTGGCTTCCGCCAGA 674
1620 tgaacaactgttccattcgtctgacaaacatgctccaaatttatacaacaacttttg 1679
675 GCACGCTGGGACATCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 734
1680 gaggttaaccaagtttgatctgaaagaagcgggaacctatggtgcgcataatcc 1739
735 GTGCTTACACGCTTCTCTTTCATCCGACAGGAC--ACCTGATTTGGCCGGCTTCT 791
1740 tgcaccccttgcataatcatatgaggtttcctgctcct 1778
792 GCCATCTCTCTGCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 830

RESULT 11
US-07-867-106-2
Sequence 2, Application US/07867106
Patent No. 5389526

GENERAL INFORMATION:
APPLICANT: Slade, Martin B
APPLICANT: Chang, Andy C M
APPLICANT: Williams, Keith L
TITLE OF INVENTION: Improved Plasmid Vectors for Cellular
TITLE OF INVENTION: Slime Moulds of the Genus Dictyostelium
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz & NO. 5389526rls
STREET: One Liberty Place 46th floor
CITY: Philadelphia
STATE: PA
COUNTRY: USA
ZIP: 19103

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/867,106
FILING DATE: 19920625

```

PRIOR APPLICATION DATA:
APPLICATION NUMBER: AU PJ 7187
APPLICATION NUMBER: PCT/AU90/00530
FILING DATE: 02-NOV-1989
ATTORNEY/AGENT INFORMATION:
NAME: Feeney, Joanne Longo
REGISTRATION NUMBER: 35,134
REFERENCE/DOCKET NUMBER: RICE-0002
TELECOMMUNICATION INFORMATION:
TELEPHONE: 215-568-3100
TELEFAX: 215-568-3439
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 5852 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ANTI-SENSE: NO
FEATURE:
NAME/KEY: CDS
LOCATION: 2378..5038
FEATURE:
NAME/KEY: CDS
LOCATION: 2378..5038
US-07-867-106-2

```

```

Query Match      1.5%; Score 44.2; DB 1; Length 5852;
Best Local Similarity 48.8%; Pred. No. 0.17;
Matches 186; Conservative 0; Mismatches 183; Indels 12; Gaps 2;

```

```

QY 586 aatgattaaattgaaacacatcttgcgtagcgaagatgattccctattctgtag 645
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 5387 ATTATATATAATGCAAGTCTCAAGATATATATATATATATATATATATATAT 5445
QY 646 gcaagtagcaaaataaaacatagaaacatctcgtactcttccctagatag 705
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 5446 TAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 5505
QY 706 atataatgtaggtt-----gctatagtagaaggtattatcgcgtccctgt 754
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 5506 ATCCGTCATGATTTTAAATAAATAAATAAATAAATAAATAAATAAATAAATA 5565
QY 755 atccctattatataaataatctttaaataatgcatcttccggtccttctgtc 814
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 5566 TTTTATTTTATTCATTAATTAATTAATTTTATTTTATTTTATTTTATTTT 5625
QY 815 tgaatttttttttgcgacactgagtagaataacattgattattatccttat 874
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 5626 TTTTATTTTAAATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 5685
QY 875 tttaagttaataaataatcgagattcgaagaaacataagaatttcttgcaga 934
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 5686 TATTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 5745
QY 935 aataaaacgaataaatga 955
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 5746 ATACAAATAAACAATAATTA 5766

```

```

RESULT 12
US-09-022-669-3
Sequence 3, Application US/09022669
Patent No. 6110717
GENERAL INFORMATION:
APPLICANT: KIRLY, KRISTINE
APPLICANT: SOUTHAN, CHRISTOPHER
APPLICANT: KNAB, ANNE
TITLE OF INVENTION: Human RCE1
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESSEE: RATNER & PRESTIA

```

```

STREET: P.O. BOX 980
CITY: VALLEY FORGE
STATE: PA
COUNTRY: USA
ZIP: 19482
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/022,669
FILING DATE: 12-FEB-1998
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 97304437.3
FILING DATE: JUNE 24, 1997
ATTORNEY/AGENT INFORMATION:
NAME: PRESTIA, PAUL F
REGISTRATION NUMBER: 23,031
REFERENCE/DOCKET NUMBER: GH-70379
TELECOMMUNICATION INFORMATION:
TELEPHONE: 610-407-0700
TELEFAX: 610-407-0701
TELEX: 846169
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 1230 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-09-022-669-3

```

```

Query Match      1.5%; Score 43.6; DB 3; Length 1230;
Best Local Similarity 53.0%; Pred. No. 0.14;
Matches 116; Conservative 0; Mismatches 100; Indels 3; Gaps 1;

```

```

QY 1560 cgcatttttttggactgcgcacgacacatgcttaagaaattacagaaggtcca 1619
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 252 CGCTCTTTTGGAGTGGCCCATTTTCCACATATATATATATATATATATATAT 311
QY 1620 tgacaactgttccatctctgtagacaacatgcttccaaatttaacaacactttg 1679
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 312 CGACGCTGGGAACAATCTTCTGTCTGCGNCGGTTCCAGTTCCCTACACAGCTGCT 371
QY 1680 gaggttaaccaagtttattcgttaagaacagcggaacotatgctgcataatcc 1739
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 372 GTGCCTACACGCTTCTCTCTTCAATCCGACAGGAC---ACCTGATTTGGCCGCTTCT 428
QY 1740 tgcattcccttgcataatcatgagggttccctgtcctt 1778
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 429 GCCATTCCTCTGCAATATACATGAGGTTCCGAGCTGTTT 467

```

```

RESULT 13
US-08-487-826B-13
Sequence 13, Application US/08487826B
Patent No. 5993827
GENERAL INFORMATION:
APPLICANT: Sim, Kim L.
APPLICANT: Chlunis, Chetan
APPLICANT: Miller, Louis H.
APPLICANT: Peterson, David S.
APPLICANT: Su, Xin-zhaun
APPLICANT: Wellens, Thomas E.
TITLE OF INVENTION: AND PLASMIDIUM FALCIPARUM ERITHROCYTE BINDING PROTEINS
NUMBER OF SEQUENCES: 45
CORRESPONDENCE ADDRESS:
ADDRESSEE: Knobbe Martens Olson & Bear
STREET: 620 Newport Center Drive 16th Floor

```

```

ZIP: 19482
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/022.669
FILING DATE: 12-FEB-1998
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 97304437.3
FILING DATE: JUNE 24, 1997
ATTORNEY/AGENT INFORMATION:
NAME: PRESTIA, PAUL F
REGISTRATION NUMBER: 23,031
REFERENCE/DOCKET NUMBER: GH-70379
TELECOMMUNICATION INFORMATION:
TELEPHONE: 610-407-0700
TELEFAX: 610-407-0701
TELEX: 846169
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1553 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-09-022-669-1

```

| | | | | | | | | | | | | | | |
|----|-----------------------|---|-----------|-------|------------|------|--------------|----|------------|------|--------|----|------|----|
| | Best local similarity | 53.0%; | Pred. No. | 0.21; | Mismatches | 116; | Conservative | 0; | Mismatches | 100; | Indels | 3; | Gaps | 1. |
| Db | 700 | CGCTTTTGTGGAGTTCCCATTTTCACCAATATTATGAGAGAGTGTTGCCGCACA | 759 | | | | | | | | | | | |
| OY | 1560 | cgccttttlttgagacttgcgacgacacacatgcttaatgagcaattacaagaagctcca | 1619 | | | | | | | | | | | |
| Db | 1620 | tgaacaacgtttccatcctgcctgacaaacatgcttcacaaatttatcacacaacactitg | 1679 | | | | | | | | | | | |
| Db | 760 | GCACGCTGGGGAAACATCTTCTGTGTCTGCTGCTGCTTCCACTGCCCTCACACAGCTGTCTTG | 819 | | | | | | | | | | | |
| OY | 1680 | gaggagttaaccaagtttgtattcgtaagaacaagcgggaacctaatgtygtgcataatcc | 1739 | | | | | | | | | | | |
| Db | 820 | GTCGCTACACAGCTTTCCTCTTCATCCACCACAGGAC---ACCTGAATGGGGCGGTTCTCT | 876 | | | | | | | | | | | |
| OY | 1740 | tgcattgcccttgcatactcatggggatttccctgthcctt | 1778 | | | | | | | | | | | |
| Db | 877 | GCAATTCCTTCTGCATTAACATGGGTTTTCCCAAGCTGTTT | 915 | | | | | | | | | | | |

RESULT 15 °
US-08-232-463-14/C
Sequence 14, Application US/08232463

APPLICANT: DORNER, F.
APPLICANT: SCHEIFLINGER, F.
APPLICANT: FALKNER, F. G.
TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 1800 Diagonal Road, Suite 500
CITY: Alexandria
STATE: VA
COUNTRY: USA
ZIP: 22313-0299
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS

THIS PAGE BLANK (1-2PTO)

PT in cancerous conditions
XX
PS Claim 15; Page 52-53; 59pp; English.
XX
XX This polypeptide comprises yeast Rce1p (Ras and a-factor converting
CC enzyme), a novel zinc metalloprotease that mediates the removal of
CC AAV tripeptide from CAAX proteins following prenylation. The yeast
CC Rce1 gene (see AAV17601) was isolated using an autoclave arrest
CC selection method. A second novel CAAX protease, Afcep1 (a-factor
CC converting enzyme), is also claimed (see AAV4301). Vectors
CC comprising the Rce1 or Afcep1 gene (see AAV17601), polypeptides encoded
CC by such vectors and recombinant cells transduced with the vectors
CC are claimed. The novel Afcep1 and Rce1p proteins can be used to
CC screen for inhibitors of prenylation which can be used in the
CC treatment of Ras dependent cancers such as colorectal and exocrine
CC pancreatic carcinomas or myeloid leukaemia. A heat shock assay for
CC assessing ras activity is also provided.
CC (N.B. this amino acid sequence of Rce1p only corresponds to the
CC translated sequence of the Rce1 gene provided in the specification
CC in the N-terminal region (amino acid residues 1-233)).
XX
SQ Sequence 315 AA;
Query Match 99.1%; Score 1636; DB 19; Length 315;
Best Local Similarity 99.4%; Pred. No. 1.1e-178;
Matches 313; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
XX
YY 1 MLOSTFLVLLXISIVLPPLATSOPEGSKRDNPRTIKSMOKLTIMLSNLFVPLFQ 60
DB 1 MLOSTFLVLLXISIVLPPLATSOPEGSKRDNPRTIKSMOKLTIMLSNLFVPLFQ 60
YY 61 SOLSSSTSHISFKDAFLGILIPGYAALPNPWOFSQFVNDLTKVAMLTLYCGPYLDF 120
DB 61 sqststshisfkdaflgilyipyaaalpnpwfsqfvdtkvcvamllylcygpyldf 120
YY 121 VLYHLNPKSSILEDFFHEPFINMSFRNFAPITTEIFTSMLTTYLNLIPHSQSLQ 180
DB 121 vlyhlmpkssiledffhefinmsfrnffapiteefitysmlltylnlphsqslsq 180
YY 181 QLFWQPSLFFGLAHANHAEOLOGSMWTVSILTTFCQILYTLFGSLTFVVRVGGN 240
DB 181 qlfwqpslffglahanhayegldqegsmwtvsillttcfqilytlfgsltfvvrvggn 240
YY 241 LMCCIILHALCNIMGFPGPSRLNLFYVDKAKGRISKLVSIWKKCYFALLVGLISLKD 300
DB 241 lwcclilhalcnimgfpgpsrlnlhftvdkkagrisklvsiwkkcyfallvlglsikd 300
YY 301 TLQTLVGTGPGYRITL 315
DB 301 tlgltvgtpgyritl 315
RESULT 2
AAV55811 standard; Protein: 314 AA.
XX
AC AAV55811;
XX
DT 29-FEB-2000 (first entry)
XX
DE Yeast RCE1 (farnesyl-directed endopeptidase) sequence.
XX
KM RCE1; farnesyl-directed endopeptidase; oncogenic ras; cancer; yeast;
XX cell proliferative disease; cellular localization; tumor.
XX
OS Saccharomyces cerevisiae.
XX
PD 02-DEC-1999;
XX
PF 19-APR-1999; 99WO-0508594.

XX
PR 22-MAY-1998; 98US-0086421.
XX
PA (ONYX-) ONYX PHARM INC.
XX
PI Chol Y, North AK, Martin GA, Bollag G;
XX
DR WPI; 2000-097123/08.
XX
PT New polypeptides, useful to identify specific modulators that inhibit
PT maturation of oncogenes such as ras, e.g. for treating cancer and other
PT cell proliferative diseases
XX
PS Disclosure; Fig 4; 31pp; English.
XX
CC The invention provides isolated RCE1 (farnesyl-directed endopeptidase)
CC polypeptides. The RCE1 polypeptides are used to identify its specific
CC modulators, i.e. agents that inhibit maturation of oncogenic ras, so are
CC potentially useful for treatment of cancer and other cell proliferative
CC diseases. Fragments of RCE1 are used to raise specific antibodies and
CC other RCE1 ligands such as nucleic acid aptamers which are used to detect
CC RCE1 for diagnostic or research purposes, including cellular localization
CC and/or distribution of the RCE1, also as therapeutic modulators and for
CC identifying RCE1-binding agents by immunoprecipitation. Nucleic acid that
CC encodes RCE1 is used: for recombinant expression of RCE1; as a source of
CC probes and primers for detecting, quantifying and isolating RCE1 e.g. for
CC diagnosis, including detecting mutations that may be associated with
CC cancers and benign tumors; as a source of antisense or ribozyme
CC therapeutics for regulating RCE1 expression; and/or as size marker. The
CC present sequence represents the yeast RCE1 amino acid sequence.
XX
SQ Sequence 314 AA;
Query Match 98.2%; Score 1621.5; DB 21; Length 314;
Best Local Similarity 98.7%; Pred. No. 4.8e-177;
Matches 311; Conservative 1; Mismatches 1; Indels 1; Gaps 1;
XX
YY 1 MLOSTFLVLLXISIVLPPLATSOPEGSKRDNPRTIKSMOKLTIMLSNLFVPLFQ 60
DB 1 MLOSTFLVLLXISIVLPPLATSOPEGSKRDNPRTIKSMOKLTIMLSNLFVPLFQ 60
YY 61 SOLSSSTSHISFKDAFLGILIPGYAALPNPWOFSQFVNDLTKVAMLTLYCGPYLDF 120
DB 61 sqststshisfkdaflgilyipyaaalpnpwfsqfvdtkvcvamllylcygpyldf 120
YY 121 VLYHLNPKSSILEDFFHEPFINMSFRNFAPITTEIFTSMLTTYLNLIPHSQSLQ 180
DB 121 vlyhlmpkssiledffhefinmsfrnffapiteefitysmlltylnlphsqslsq 179
YY 181 QLFWQPSLFFGLAHANHAEOLOGSMWTVSILTTFCQILYTLFGSLTFVVRVGGN 240
DB 181 qlfwqpslffglahanhayegldqegsmwtvsillttcfqilytlfgsltfvvrvggn 239
YY 241 LMCCIILHALCNIMGFPGPSRLNLFYVDKAKGRISKLVSIWKKCYFALLVGLISLKD 300
DB 241 lwcclilhalcnimgfpgpsrlnlhftvdkkagrisklvsiwkkcyfallvlglsikd 299
YY 301 TLQTLVGTGPGYRITL 315
DB 301 tlgltvgtpgyritl 314
RESULT 3
AAV55810 standard; Protein: 329 AA.
XX
AC AAV55810;
XX
DT 29-FEB-2000 (first entry)
XX
DE Mouse RCE1 (farnesyl-directed endopeptidase) sequence.
XX

KW RCE1; farnesyl-directed endopeptidase; oncogenic ras; cancer;
 XX cell proliferative disease; cellular localization; tumor.
 XX Mus sp.
 PN WO9961628-A2.
 XX 02-DEC-1999.
 PD 02-DEC-1999.
 XX 19-APR-1999; 99WO-US08594.
 XX 22-MAY-1998; 98US-0086421.
 XX (ONYX-) ONYX PHARM INC.
 PI Choi Y, North AK, Martin GA, Bollag G;
 DR WPI; 2000-097123/08.
 DR N-PSDB; AA239729.
 XX New polypeptides, useful to identify specific modulators that inhibit
 PT maturation of oncogenes such as ras, e.g. for treating cancer and other
 PT cell proliferative diseases
 PS Claim 8; Fig 3; 31pp; English.
 XX The invention provides isolated RCE1 (farnesyl-directed endopeptidase)
 CC polypeptides. The RCE1 polypeptides are used to identify its specific
 CC modulators, i.e. agents that inhibit maturation of oncogenic ras, so are
 CC potentially useful for treatment of cancer and other cell proliferative
 CC diseases. Fragments of RCE1 are used to raise specific antibodies and
 CC other RCE1 ligands such as nucleic acid aptamers which are used to detect
 CC RCE1 for diagnostic or research purposes, including cellular localization
 CC and/or distribution of the RCE1, also as therapeutic modulators and for
 CC identifying RCE1-binding agents by immunoprecipitation. Nucleic acid that
 CC encodes RCE1 is used; for recombinant expression of RCE1; as a source of
 CC probes and primers for detecting, quantifying and isolating RCE1 e.g. for
 CC diagnosis, including detecting mutations that may be associated with
 CC cancers and benign tumors; as a source of antisense or ribozyme
 CC therapeutics for regulating RCE1 expression; and/or as size marker. The
 CC present sequence represents the mouse RCE1 amino acid sequence.
 XX SO Sequence 329 AA;

Query Match 14.1%; Score 233.5; DB 21; Length 329;
 Best Local Similarity 30.9%; Pred. No. 4.1e-18;
 Matches 86; Conservative 41; Mismatches 80; Indels 71; Gaps 16;

QY 4 FSTFLVLLTISIVPLVLTATSOEGSKRDNPRTIKRMOKLTIMLSNF-LVPLFQSQ 62
 DB 40 fscf-----slacsvvgslwykwse--lprdhpaikrr--stslvsslsplcvllwre 91
 QY 63 LSSST---TSHISFKDAFLGIIIPGYAALPNPWFQSFVKDLKVCAMLT--LYCGPV 117
 DB 92 ltlgqptsltlm-gfrlglfpa--aal-----lalltmllflgpl 131
 QY 118 LDFV-----LYHLNPKSSILEDYHEFLNWS-----FRNFIPADITEIF 159
 DB 132 mqslmdpcdltdgllkvlafrs-----warcldmrwlnqvlapleelv 178
 QY 160 YTSMLLTLYNLTPHSGLSQGLFWQPSLFFGLAHAAHNAVEQOEGSMITVSLTLTFCQ 219
 DB 179 firaampm---lapcglg-pavltcp-lifgvaahfhlleqltrfgrsvgsiltvsaaqf 233
 QY 220 ILVTTFLGGLKRVFVTRGCLMCCILHALLCNIMGFP 257
 DB 234 fsytavfgaytafltrt-ghllgprlvchsfcmymgfp 270

RESULT 4
 AAW86010
 ID AAW86010 standard; Protein; 329 AA.

XX AC AAW86010;
 XX DT 29-MAR-1999 (first entry)
 XX DE Mouse CAAX processing enzyme RCE1 homologue mrce1p.
 XX DE RCE1; mrce1p; CAAX processing enzyme; mouse; tumour; cancer;
 XX KM therapy; diagnosis; Ras protein; endoprotease.
 XX Mus sp.
 OS WO9854333-A2.
 PN 03-DEC-1998.
 PD 03-DEC-1998.
 XX 02-JUN-1998; 98WO-US11415.
 XX 14-JUL-1997; 97US-0052389.
 PR 02-JUN-1997; 97US-0047369.
 XX (ACAC-) ACACIA BIOSCIENCES INC.
 PA Ashby MN, Dimster-Denk DG, Phillips JW;
 PI WPI; 1999-059843/05.
 DR N-PSDB; AAW80323-24.
 XX New DNA encoding mammalian CAAX-processing enzymes - used e.g. to
 PT treat CAAX-protein mediated diseases such as cancers and tumours
 PT associated with mutant Ras
 PS Claim 15; Fig 2B; 98pp; English.
 XX This is the amino acid sequence of mrce1p, a murine functional
 CC homologue of the yeast prenylation-dependent CAAX endoprotease
 CC Rcep1 that contributes to the processing of a-factor and the yeast
 CC Ras protein. This mammalian homologue represents a potential
 CC target to block the oncogenic action of mutant Ras protein in
 CC tumours or, more generally, to modulate the activity of prenylated
 CC peripheral membrane proteins. The mrce1p amino acid sequence was
 CC deduced from the nucleotide sequence of isolated cDNA and genomic
 CC clones (see AAW80323-24). Host cells transformed with mammalian
 CC CAAX processing enzyme DNA (see AAW80322-25) can be used to produce
 CC recombinant polypeptides (see AAW86009-12) used for in vitro
 CC screening of inhibitors and to raise antibodies. The inhibitors
 CC are used to treat CAAX-protein mediated diseases, especially
 CC cancers and tumours associated with abnormal Ras activity.
 CC Antibodies are used to screen for expression of CAAX processing
 CC proteins, for affinity purification and in immunoassays to
 CC determine levels of CAAX processing proteins or their subcellular
 CC localization and to confirm interaction with candidate binding
 CC proteins.
 XX SO Sequence 329 AA;

Query Match 14.0%; Score 230.5; DB 20; Length 329;
 Best Local Similarity 30.6%; Pred. No. 9.1e-18;
 Matches 85; Conservative 40; Mismatches 82; Indels 71; Gaps 16;

QY 4 FSTFLVLLTISIVPLVLTATSOEGSKRDNPRTIKRMOKLTIMLSNF-LVPLFQSQ 62
 DB 40 fscf-----slacsvvgslwykwse--lprdhpaikrr--stslvsslsplcvllwre 91
 QY 63 LSSST---TSHISFKDAFLGIIIPGYAALPNPWFQSFVKDLKVCAMLT--LYCGPV 117
 DB 92 ltlgqptsltlm-gfrlglfpa--allp-----lilltmllflgpl 131
 QY 118 LDFV-----LYHLNPKSSILEDYHEFLNWS-----FRNFIPADITEIF 159
 DB 132 mqslmdpcdltdgllkvlafrs-----warcldmrwlnqvlapleelv 178

Sequence 293 AA;

diagnosis and treatment of cancers.

diagnosis and treatment of cancers.

AC AA003600;
 XX 24-OCT-2001 (first entry)
 DT
 DE Human ras converting endoprotease (RCE).
 XX
 XX Human; ras converting endoprotease; RCE inhibitor; placental;
 KW prenyl cysteine specific carboxymethyltransferase; cancer.
 XX
 XX Homo sapiens.
 OS
 XX US6261793-B1.
 PN
 XX 17-JUL-2001.
 PD
 XX 04-MAR-1999; 99US-0262749.
 PF
 XX 04-MAR-1999; 99US-0262749.
 PR
 XX (SCHE) SCHERING CORP.
 PA
 XX Whyte D, McGuirk M, Nunez-Oliva I, Hockenberry T, Pai J;
 PI
 XX WPI: 2001-450487/48.
 DR N-PSDB; AAS04152.
 XX
 PT Identifying human Ras Converting Endoprotease (RCE) inhibitors by
 PT contacting radiolabelled prenylated protein with sample to be tested for
 PT presence of inhibitor, and RCE, and measuring amount of labeled
 PT tripeptide released -
 XX
 PS Claim 1; Column 15-18; 10pp; English.
 XX
 CC The present sequence representing human ras converting endoprotease
 CC (RCE) is isolated from a human placental cDNA library. RCE is used to
 CC identify inhibitors of human RCE. Methods of identifying RCE inhibitors
 CC comprises contacting a radiolabelled prenylated protein with the sample
 CC to be tested for RCE inhibitors and RCE, and measuring the amount of
 CC labelled tripeptide released; or contacting unlabelled prenylated
 CC protein with the sample, RCE, prenyl cysteine specific
 CC carboxymethyltransferase and, radiolabelled methyl group donor (MD),
 CC and measuring MD incorporated into the prenylated protein. The methods
 CC of the invention are useful for identifying RCE inhibitors, which are
 CC useful for treating cancer.
 XX
 SQ Sequence 329 AA;

```

Query Match 13.8%: Score 227.5; DB 22; Length 329;
Best Local Similarity 28.7%: Pred. No. 2e-17; Indels 89; Gaps 14;
Matches 81; Conservative 39; Mismatches 73;

QY 11 LVYSISVYPLVATSQPESSKRDNPRTIKSRQKLTIMLSNF-LVPELOSSTSTSH 69
   | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 43 lslacgygsllyvwase--lprdhavkrrf--tsvlyvsslsplcvllwelf----- 93

QY 70 ISFKDAFLGIGIIPG-----YVALPNPWFQSFVKDKLCVAMLLT-LY 113
   || || | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 94 -----g|qpgysl|tlmgf|leg|flaal-----lp|llm|l|f 127

QY 114 CGPVLDV-----LYHLNPKSSILEDYHEFLNWS-----FRNFIPAPIT 155
   ||||| | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 128 lqplmglsmdcpodladgkvvlaps-----warc|tdm|w|l|rqv|ap|l 174

QY 156 EEIIFTSMLLTYYLNIPIHSQSLQGLFWQPSLFGCLAHNAHAYEDQESMTYSIILT 215
   || : : : : | | | | | | | | | | | | | | | | | | | | | | | |
Db 175 eelivfacmlpm--k|apcm|g-l-pav|t|cp--l|f|vaf|h|h|l|e|g|l|f|f|s|v|g|n|f|s 229

QY 216 TCGQIILYTLTFGLTFKFEVFPRTGNCIMCCIIHALCNIGFP 257
   ||||| | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 230 aafqfsytavf|g|y|t|a|f|l|f|t--g|h|l|g|p|l|s|f|c|n|g|f|p 270

```

| | |
|--------|---|
| RESULT | 11 |
| ID | AAy87337 |
| XX | AAy87337 standard; Protein; 225 AA. |
| AC | |
| XX | AAy87337; |
| DT | |
| XX | 11-MAY-2000 (first entry) |
| DE | |
| XX | Human signal peptide containing protein HSP-114 SEQ ID NO:114. |
| KM | Human; signal peptide-containing protein; HSP; diagnosis: cancer; |
| KM | inflammation; cardiovascular disease; anticancer; anti-inflammatory; |
| KM | antimicrobial; neotrophic; neuroprotective; cardiovascular; hepatotropic; |
| KM | antisthmatic; gene therapy; cell proliferation; neurological disorder; |
| KM | reproductive disorder; developmental disorder; arteriosclerosis; |
| KM | cirrhosis; psoriasis; acquired immune deficiency syndrome; anaemia; |
| KM | asthma; Crohn's disease; infection; Alzheimer's disease; schizophrenia; |
| KM | Parkinson's disease; Huntington's diseases; ovulatory defect; |
| OS | |
| XX | Homo sapiens. |
| PN | |
| XX | WO200000610-A2. |
| PD | |
| XX | 06-JAN-2000. |
| PF | |
| XX | 25-JUN-1999; 99WO-US14484. |
| PR | |
| XX | 26-JUN-1998; 98US-0090762. |
| PR | 31-JUL-1998; 98US-0094983. |
| PR | 01-OCT-1998; 98US-0102686. |
| PR | 11-DEC-1998; 98US-0112129. |
| XX | |
| PA | (INCY-) INCYTE PHARM INC. |
| PI | |
| XX | Lal P, Tang YT, Gorgone GA, Corley NC, Guegler KJ, Baughn MR; |
| PI | Akeblom IE, Au-young J, Yue H, Patterson C, Reddy R, Hiltman JL; |
| PI | Bandman O; |
| XX | |
| DR | WPI: 2000-160673/14. |
| XX | |
| DR | N-PSDB; AAy87337. |
| XX | |
| PT | New human signal peptide-containing proteins useful in treatment, |
| PT | prevention and diagnosis of e.g. cancer, inflammation and |
| PT | cardiovascular disease |
| PS | |
| XX | Claim 1; Page 233; 327pp; English. |
| XX | |
| XX | AAy87337 to AAy87337 encode AAy87337 which represent the |
| CC | human signal peptide-containing proteins HSP-1 to HSP-134. HSPs have |
| CC | anticancer, anti-inflammatory, antimicrobial, neotrophic, hepatotropic, |
| CC | neuroprotective, cardiovascular and antisthmatic activities, and can |
| CC | be used in gene therapy. HSPs can be used to treat or prevent disorders |
| CC | associated with decreased activity or function of HSP. Antagonists of |
| CC | HSP are used to treat or prevent disorders associated with increased |
| CC | activity or function of HSP. Such diseases include cell proliferation |
| CC | (including cancer), inflammation, cardiovascular, neurological, |
| CC | reproductive or developmental disorders, (e.g. arteriosclerosis, |
| CC | cirrhosis, psoriasis, acquired immune deficiency syndrome, anaemia, |
| CC | asthma, Crohn's disease, microbial or other infections, congestive or |
| CC | ischemic heart disease, Alzheimer's, Parkinson's or Huntington's |
| CC | diseases, schizophrenia, ovulatory defects, muscular dystrophy). HSP |
| CC | nucleic acids can be used for the recombinant production of HSP, for |
| CC | detecting HSP in standard hybridisation and amplification assays (for |
| CC | diagnosis and monitoring), in gene therapy, as antisense, |
| CC | triplex-forming or ribozyme therapeutics, for detecting related sequences |
| CC | or genetic variations, and for chromosomal mapping. HSP are also used to |
| CC | raise specific antibodies (Ab) and to screen for agonists and |
| CC | antagonists (potential therapeutic agents). Ab are used to diagnose, or |
| CC | monitor, HSP-related diseases (in usual immunoassays), as therapeutic |
| CC | antagonists, in competitive drug screens, and for purification of HSP |
| XX | from natural sources. |

SQ Sequence 225 AA;

Query Match 12.4% Score 205; DB 21; Length 225;

Best Local Similarity 31.8%; Pred. No. 4.5e-15;

Matches 63; Conservative 26; Mismatches 51; Indels 58; Gaps 10;

QY 80 GIRGYVALPMPQSFQKDLTKVCAMLLT--LYGCPVDFV-----LYHLLN 127

Db 7 glfpa--allp-----llltmllfigplnqlsmdecdladgkxvla 47

QY 128 PKSSILEDYFHEFLNIMS-----FNNFIAPITEIEFYTSMLTTYLNLRHSQLSY 179

Db 48 prs-----warcltmrwlrngvlaplteelvfrcacmlpm---lapomgig- 90

QY 180 QQLFWQPSLEFGLAHNAHNAVQLOEGSMITVSIILTCFQLYLTTLFEGLRKPFVRRGG 239

Db 91 pavltcp-lffgvhfhfhllqqlfrgssvgnlftsaafqtsyavfayataflirt-g 148

QY 240 NLWCCIIHALCNIMGFP 257

Db 149 hllgprvlchscfcmymgfp 166

RESULT 12

AAC30791 ID AAC30791 standard; Protein: 461 AA.

XX AAC30791;

DT 17-OCT-2000 (first entry)

XX Arebidopsis thaliana protein fragment SEQ ID NO: 36874.

KW Protein identification: signal transduction pathway; metabolic pathway;

KW hybridisation assay; genetic mapping; gene expression control; promoter;

XX termination sequence.

XX Arebidopsis thaliana.

XX EPI033405-A2.

PD 06-SEP-2000.

PF 25-FEB-2000; 2000EP-0301439.

PR 25-FEB-1999; 99US-0121825.

PR 05-MAR-1999; 99US-0123180.

PR 09-MAR-1999; 99US-0123548.

PR 23-MAR-1999; 99US-0125788.

PR 25-MAR-1999; 99US-0126264.

PR 29-MAR-1999; 99US-0126785.

PR 01-APR-1999; 99US-0127462.

PR 06-APR-1999; 99US-0128234.

PR 08-APR-1999; 99US-0128714.

PR 16-APR-1999; 99US-0129845.

PR 19-APR-1999; 99US-0130047.

PR 21-APR-1999; 99US-0130449.

PR 23-APR-1999; 99US-0130510.

PR 28-APR-1999; 99US-0130891.

PR 30-APR-1999; 99US-0131449.

PR 04-MAY-1999; 99US-0132407.

PR 05-MAY-1999; 99US-0132484.

PR 06-MAY-1999; 99US-0132485.

PR 06-MAY-1999; 99US-0132486.

PR 07-MAY-1999; 99US-0132487.

PR 11-MAY-1999; 99US-0134256.

PR 14-MAY-1999; 99US-0134218.

PR 14-MAY-1999; 99US-0134219.

PR 14-MAY-1999; 99US-0134221.

PR 14-MAY-1999; 99US-0134370.

PR 18-MAY-1999; 99US-0134768.

PR 19-MAY-1999; 99US-0134941.

PR 20-MAY-1999; 99US-0135124.

PR 21-MAY-1999; 99US-0135353.

PR 24-MAY-1999; 99US-0135629.

PR 25-MAY-1999; 99US-0136021.

PR 27-MAY-1999; 99US-0136392.

PR 28-MAY-1999; 99US-0136782.

PR 01-JUN-1999; 99US-0137222.

PR 03-JUN-1999; 99US-0137528.

PR 04-JUN-1999; 99US-0137502.

PR 07-JUN-1999; 99US-0137724.

PR 08-JUN-1999; 99US-0138074.

PR 10-JUN-1999; 99US-0138540.

PR 10-JUN-1999; 99US-0138847.

PR 14-JUN-1999; 99US-0139119.

PR 16-JUN-1999; 99US-0139452.

PR 16-JUN-1999; 99US-0139453.

PR 17-JUN-1999; 99US-0139492.

PR 18-JUN-1999; 99US-0139454.

PR 18-JUN-1999; 99US-0139455.

PR 18-JUN-1999; 99US-0139456.

PR 18-JUN-1999; 99US-0139457.

PR 18-JUN-1999; 99US-0139458.

PR 18-JUN-1999; 99US-0139459.

PR 18-JUN-1999; 99US-0139460.

PR 18-JUN-1999; 99US-0139461.

PR 18-JUN-1999; 99US-0139462.

PR 18-JUN-1999; 99US-0139463.

PR 18-JUN-1999; 99US-0139750.

PR 18-JUN-1999; 99US-0139763.

PR 21-JUN-1999; 99US-0139817.

PR 22-JUN-1999; 99US-0139859.

PR 23-JUN-1999; 99US-0140353.

PR 23-JUN-1999; 99US-0140354.

PR 24-JUN-1999; 99US-0140695.

PR 28-JUN-1999; 99US-0140823.

PR 29-JUN-1999; 99US-0140991.

PR 30-JUN-1999; 99US-0141287.

PR 01-JUL-1999; 99US-0141842.

PR 01-JUL-1999; 99US-0142154.

PR 02-JUL-1999; 99US-0142055.

PR 06-JUL-1999; 99US-0142390.

PR 08-JUL-1999; 99US-0142803.

PR 09-JUL-1999; 99US-0142970.

PR 12-JUL-1999; 99US-0142977.

PR 13-JUL-1999; 99US-0143642.

PR 14-JUL-1999; 99US-0143624.

PR 15-JUL-1999; 99US-0144005.

PR 16-JUL-1999; 99US-0144085.

PR 16-JUL-1999; 99US-0144086.

PR 19-JUL-1999; 99US-0144325.

PR 19-JUL-1999; 99US-0144331.

PR 19-JUL-1999; 99US-0144332.

PR 19-JUL-1999; 99US-0144333.

PR 19-JUL-1999; 99US-0144334.

PR 19-JUL-1999; 99US-0144335.

PR 20-JUL-1999; 99US-0144352.

PR 20-JUL-1999; 99US-0144632.

PR 20-JUL-1999; 99US-0144884.

PR 21-JUL-1999; 99US-0144814.

PR 21-JUL-1999; 99US-0145086.

PR 21-JUL-1999; 99US-0145088.

PR 22-JUL-1999; 99US-0145085.

PR 22-JUL-1999; 99US-0145087.

PR 22-JUL-1999; 99US-0145089.

PR 22-JUL-1999; 99US-0145192.

PR 23-JUL-1999; 99US-0145145.

PR 23-JUL-1999; 99US-0145218.

PR 26-JUL-1999; 99US-0145224.

PR 26-JUL-1999; 99US-0145276.

PR 27-JUL-1999; 99US-0145913.

PR 27-JUL-1999; 99US-0145918.

PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0148171.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151086.
PR 30-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158222.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 18-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.

PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161922.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 5.94; Score 97.5; DB 21; Length 461;
Best Local Similarity 22.18; Pred. No. 0.024;
Matches 44; Conservative 30; Mismatches 86; Indels 39; Gaps 7;

QY 70 ISFKDAFLGCIIPGYAALPNPQFSQFVD-----LTKCYAMLLTTCGPVLPVL 122
||| ||| : : : : : ||| : : : : :
Db 277 isfpallgtyvgrnwsqapnncrvktipripckkwyilpsvslm----g9llpf-- 330

QY 123 YHLLNPKRSILEDYHFEFLNWSFRNFEPADITEEFYTSMLTTYLNL- PHSQLSYQ 181
||| ||| : : : : : ||| : : : : :
Db 331 -----gsafiemyfvtfsfemvkyyvymfllvflvivecvltvgtyflnaen 383

QY 182 LFMQPSLFFGLAHNAHAYEOLQEGSMITVSILTTCTOILTYLPGL--TKFVFVRIG 239
||| ||| : : : : : ||| : : : : :
Db 384 yhwqwtstfsaa-----stavyvylsylyyyvvtkmqsfqfsltyf---gy 427

QY 240 NLWCCIIHALCNMGFPG 258
: ||| ||| : : : : :
Db 428 tnmfcldgltgagayyig 446.

RESULT 13
AAG45639
ID AAG45639 standard; Protein: 490 AA.
XX AC AAG45639;
XX DT 18-OCT-2000 (first entry)
XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 57322.
XX KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX OS Arabidopsis thaliana.
XX XX
XX PN EP1033405-A2.
XX PD 06-SEP-2000.
XX PF 25-FEB-2000; 2000EP-0301439.
XX XX
XX 25-FEB-1999; 99US-0121825.
XX 05-MAR-1999; 99US-0123180.
XX 09-MAR-1999; 99US-0123548.
XX 23-MAR-1999; 99US-0125788.
XX 25-MAR-1999; 99US-0126264.
XX 29-MAR-1999; 99US-0126785.
XX 01-APR-1999; 99US-0127462.
XX 06-APR-1999; 99US-0128234.
XX 08-APR-1999; 99US-0128714.
XX 16-APR-1999; 99US-0129845.
XX 19-APR-1999; 99US-0130077.
XX 21-APR-1999; 99US-0130449.
XX 23-APR-1999; 99US-0130510.
XX 23-APR-1999; 99US-0130891.
XX 28-APR-1999; 99US-0131449.
XX 30-APR-1999; 99US-0132048.
XX 30-APR-1999; 99US-0132407.
XX 04-MAY-1999; 99US-0132484.
XX 05-MAY-1999; 99US-0132485.
XX 06-MAY-1999; 99US-0132486.
XX 07-MAY-1999; 99US-0132487.
XX 07-MAY-1999; 99US-0132863.

PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137752.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140635.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142380.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.

PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147250.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148311.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149902.
PR 25-AUG-1999; 99US-0150566.
PR 25-AUG-1999; 99US-0150884.
PR 26-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156556.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157763.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.

| Query Match | 5.9% | Score 97.5 | DB 21 | Length 490 |
|-----------------------|--|---|-------|----------------------------------|
| Best Local Similarity | 22.1% | Pred. No. 0.026 | | |
| Matches | 44 | Conservative | 30 | Mismatches 86; Indels 39; Gaps 7 |
| QY | 70 | ISFNKAFGLGIGIPFYVALPMPQFSGVKD-----LTFGVAMLLTYLGGPVLDFVL | 122 | |
| Db | 306 | ISFIALIGLVGVGWSGAPNCPVKTKIPRIPEKKWYLPSVSLM--GGLIPF-- | 359 | |
| QY | 123 | YHLNPKRSSILEDYFHEFNIMSPFNFIAPRTIEEFYTSMLTYLNLI-PSOLSVOQ | 181 | |
| Db | 360 | -----GSIFIEMYFVFSTSFNPKVYVYVYGFMLLVFVLYLVLCVTLVGTGFLINAE | 412 | |
| QY | 182 | LFWQPSLEFFGLAHNAHNAVEQLOEGSMVTYVSIILTTQFQILYTLFGGL-TRFVFNRTGG | 239 | |
| Db | 413 | yhwqstffssaa-----stavyvlysllyyyvktkmsgffqtsfzf--gy | 456 | |
| QY | 240 | NLMCCIIILHALCNIMGFPG | 258 | |
| Db | 457 | tmnfclgldgllcgvaylyg | 475 | |
| RESULT 14 | | | | |
| AAAG21346 | | | | |
| ID | AAAG21346 standard; Protein: 496 AA. | | | |
| XX | AAAG21346; | | | |
| XX | 17-OCT-2000 (first entry) | | | |
| DE | Arabidopsis thaliana protein fragment SEQ ID NO: 23871. | | | |
| XX | Protein identification; signal transduction pathway; metabolic pathway; | | | |
| KW | hybridisation assay; genetic mapping; gene expression control; promoter; | | | |
| KW | termination sequence. | | | |
| OS | Arabidopsis thaliana. | | | |
| PN | EP1033405-A2. | | | |
| PD | 06-SEP-2000. | | | |
| PE | 25-FEB-2000; 2000EP-0301439. | | | |
| PR | 25-FEB-1999; | 99US-0121825. | | |
| PR | 09-MAR-1999; | 98US-0123180. | | |
| PR | 23-MAR-1999; | 99US-0123548. | | |
| PR | 23-MAR-1999; | 99US-0125788. | | |
| PR | 29-MAR-1999; | 99US-0126264. | | |
| PR | 01-APR-1999; | 99US-0126785. | | |
| PR | 06-APR-1999; | 99US-0127462. | | |
| PR | 08-APR-1999; | 99US-0128234. | | |
| PR | 16-APR-1999; | 99US-0128714. | | |
| PR | 19-APR-1999; | 99US-0128845. | | |
| PR | 21-APR-1999; | 99US-0130077. | | |
| PR | 23-APR-1999; | 99US-0130449. | | |
| PR | 23-APR-1999; | 99US-0130510. | | |
| PR | 28-APR-1999; | 99US-0130891. | | |
| PR | 30-APR-1999; | 99US-0131449. | | |
| PR | 30-APR-1999; | 99US-0132048. | | |
| PR | 30-APR-1999; | 99US-0132407. | | |

| | | |
|----|-------------|--------------|
| PR | 04-MAY-1999 | 9905-0132484 |
| PR | 05-MAY-1999 | 9905-0132486 |
| PR | 06-MAY-1999 | 9905-0132487 |
| PR | 07-MAY-1999 | 9905-0132488 |
| PR | 08-MAY-1999 | 9905-0132489 |
| PR | 09-MAY-1999 | 9905-0132490 |
| PR | 10-MAY-1999 | 9905-0132491 |
| PR | 11-MAY-1999 | 9905-0132492 |
| PR | 12-MAY-1999 | 9905-0132493 |
| PR | 13-MAY-1999 | 9905-0132494 |
| PR | 14-MAY-1999 | 9905-0132495 |
| PR | 15-MAY-1999 | 9905-0132496 |
| PR | 16-MAY-1999 | 9905-0132497 |
| PR | 17-MAY-1999 | 9905-0132498 |
| PR | 18-MAY-1999 | 9905-0132499 |
| PR | 19-MAY-1999 | 9905-0132500 |
| PR | 20-MAY-1999 | 9905-0132501 |
| PR | 21-MAY-1999 | 9905-0132502 |
| PR | 22-MAY-1999 | 9905-0132503 |
| PR | 23-MAY-1999 | 9905-0132504 |
| PR | 24-MAY-1999 | 9905-0132505 |
| PR | 25-MAY-1999 | 9905-0132506 |
| PR | 26-MAY-1999 | 9905-0132507 |
| PR | 27-MAY-1999 | 9905-0132508 |
| PR | 28-MAY-1999 | 9905-0132509 |
| PR | 29-MAY-1999 | 9905-0132510 |
| PR | 30-MAY-1999 | 9905-0132511 |
| PR | 31-MAY-1999 | 9905-0132512 |
| PR | 01-JUN-1999 | 9905-0132513 |
| PR | 02-JUN-1999 | 9905-0132514 |
| PR | 03-JUN-1999 | 9905-0132515 |
| PR | 04-JUN-1999 | 9905-0132516 |
| PR | 05-JUN-1999 | 9905-0132517 |
| PR | 06-JUN-1999 | 9905-0132518 |
| PR | 07-JUN-1999 | 9905-0132519 |
| PR | 08-JUN-1999 | 9905-0132520 |
| PR | 09-JUN-1999 | 9905-0132521 |
| PR | 10-JUN-1999 | 9905-0132522 |
| PR | 11-JUN-1999 | 9905-0132523 |
| PR | 12-JUN-1999 | 9905-0132524 |
| PR | 13-JUN-1999 | 9905-0132525 |
| PR | 14-JUN-1999 | 9905-0132526 |
| PR | 15-JUN-1999 | 9905-0132527 |
| PR | 16-JUN-1999 | 9905-0132528 |
| PR | 17-JUN-1999 | 9905-0132529 |
| PR | 18-JUN-1999 | 9905-0132530 |
| PR | 19-JUN-1999 | 9905-0132531 |
| PR | 20-JUN-1999 | 9905-0132532 |
| PR | 21-JUN-1999 | 9905-0132533 |
| PR | 22-JUN-1999 | 9905-0132534 |
| PR | 23-JUN-1999 | 9905-0132535 |
| PR | 24-JUN-1999 | 9905-0132536 |
| PR | 25-JUN-1999 | 9905-0132537 |
| PR | 26-JUN-1999 | 9905-0132538 |
| PR | 27-JUN-1999 | 9905-0132539 |
| PR | 28-JUN-1999 | 9905-0132540 |
| PR | 29-JUN-1999 | 9905-0132541 |
| PR | 30-JUN-1999 | 9905-0132542 |
| PR | 01-JUL-1999 | 9905-0132543 |
| PR | 02-JUL-1999 | 9905-0132544 |
| PR | 03-JUL-1999 | 9905-0132545 |
| PR | 04-JUL-1999 | 9905-0132546 |
| PR | 05-JUL-1999 | 9905-0132547 |
| PR | 06-JUL-1999 | 9905-0132548 |
| PR | 07-JUL-1999 | 9905-0132549 |
| PR | 08-JUL-1999 | 9905-0132550 |
| PR | 09-JUL-1999 | 9905-0132551 |
| PR | 10-JUL-1999 | 9905-0132552 |
| PR | 11-JUL-1999 | 9905-0132553 |
| PR | 12-JUL-1999 | 9905-0132554 |
| PR | 13-JUL-1999 | 9905-0132555 |
| PR | 14-JUL-1999 | 9905-0132556 |
| PR | 15-JUL-1999 | 9905-0132557 |
| PR | 16-JUL-1999 | 9905-0132558 |
| PR | 17-JUL-1999 | 9905-0132559 |
| PR | 18-JUL-1999 | 9905-0132560 |
| PR | 19-JUL-1999 | 9905-0132561 |
| PR | 20-JUL-1999 | 9905-0132562 |
| PR | 21-JUL-1999 | 9905-0132563 |
| PR | 22-JUL-1999 | 9905-0132564 |
| PR | 23-JUL-1999 | 9905-0132565 |
| PR | 24-JUL-1999 | 9905-0132566 |
| PR | 25-JUL-1999 | 9905-0132567 |
| PR | 26-JUL-1999 | 9905-0132568 |
| PR | 27-JUL-1999 | 9905-0132569 |
| PR | 28-JUL-1999 | 9905-0132570 |
| PR | 29-JUL-1999 | 9905-0132571 |
| PR | 30-JUL-1999 | 9905-0132572 |
| PR | 31-JUL-1999 | 9905-0132573 |
| PR | 01-AUG-1999 | 9905-0132574 |
| PR | 02-AUG-1999 | 9905-0132575 |
| PR | 03-AUG-1999 | 9905-0132576 |
| PR | 04-AUG-1999 | 9905-0132577 |
| PR | 05-AUG-1999 | 9905-0132578 |
| PR | 06-AUG-1999 | 9905-0132579 |
| PR | 07-AUG-1999 | 9905-0132580 |
| PR | 08-AUG-1999 | 9905-0132581 |
| PR | 09-AUG-1999 | |

PR 23-APR-1999; 990S-0130510.
PR 28-APR-1999; 990S-0130891.
PR 28-APR-1999; 990S-0131449.
PR 30-APR-1999; 990S-0132048.
PR 30-APR-1999; 990S-0132407.
PR 04-MAY-1999; 990S-0132484.
PR 05-MAY-1999; 990S-0132485.
PR 06-MAY-1999; 990S-0132486.
PR 06-MAY-1999; 990S-0132487.
PR 07-MAY-1999; 990S-0132483.
PR 11-MAY-1999; 990S-0134256.
PR 14-MAY-1999; 990S-0134218.
PR 14-MAY-1999; 990S-0134219.
PR 14-MAY-1999; 990S-0134370.
PR 18-MAY-1999; 990S-0134376.
PR 19-MAY-1999; 990S-0134941.
PR 20-MAY-1999; 990S-0135124.
PR 21-MAY-1999; 990S-0135353.
PR 24-MAY-1999; 990S-0135629.
PR 25-MAY-1999; 990S-0136021.
PR 27-MAY-1999; 990S-0136392.
PR 28-MAY-1999; 990S-0136782.
PR 01-JUN-1999; 990S-0137222.
PR 03-JUN-1999; 990S-0137528.
PR 04-JUN-1999; 990S-0137502.
PR 07-JUN-1999; 990S-0137724.
PR 08-JUN-1999; 990S-0138094.
PR 10-JUN-1999; 990S-0138540.
PR 10-JUN-1999; 990S-0138847.
PR 14-JUN-1999; 990S-0139119.
PR 16-JUN-1999; 990S-0139452.
PR 17-JUN-1999; 990S-0139453.
PR 18-JUN-1999; 990S-0139492.
PR 18-JUN-1999; 990S-0139454.
PR 18-JUN-1999; 990S-0139455.
PR 18-JUN-1999; 990S-0139456.
PR 18-JUN-1999; 990S-0139457.
PR 18-JUN-1999; 990S-0139458.
PR 18-JUN-1999; 990S-0139459.
PR 18-JUN-1999; 990S-0139460.
PR 18-JUN-1999; 990S-0139461.
PR 18-JUN-1999; 990S-0139462.
PR 18-JUN-1999; 990S-0139463.
PR 18-JUN-1999; 990S-0139750.
PR 18-JUN-1999; 990S-0139763.
PR 21-JUN-1999; 990S-0139817.
PR 22-JUN-1999; 990S-0139899.
PR 23-JUN-1999; 990S-0140353.
PR 23-JUN-1999; 990S-0140354.
PR 24-JUN-1999; 990S-0140695.
PR 28-JUN-1999; 990S-0140823.
PR 28-JUN-1999; 990S-0140991.
PR 30-JUN-1999; 990S-0141287.
PR 01-JUL-1999; 990S-0141842.
PR 02-JUL-1999; 990S-0142055.
PR 06-JUL-1999; 990S-0142390.
PR 08-JUL-1999; 990S-0142803.
PR 09-JUL-1999; 990S-0142920.
PR 12-JUL-1999; 990S-0143542.
PR 13-JUL-1999; 990S-0143542.
PR 14-JUL-1999; 990S-0143624.
PR 15-JUL-1999; 990S-0144005.
PR 16-JUL-1999; 990S-0144085.
PR 16-JUL-1999; 990S-0144086.
PR 19-JUL-1999; 990S-0144325.
PR 19-JUL-1999; 990S-0144331.
PR 19-JUL-1999; 990S-0144332.
PR 19-JUL-1999; 990S-0144333.
PR 19-JUL-1999; 990S-0144334.
PR 19-JUL-1999; 990S-0144335.
PR 20-JUL-1999; 990S-0144352.

PR 20-JUL-1999; 990S-0144632.
PR 20-JUL-1999; 990S-0144884.
PR 21-JUL-1999; 990S-0144814.
PR 21-JUL-1999; 990S-0145086.
PR 21-JUL-1999; 990S-0145088.
PR 22-JUL-1999; 990S-0145085.
PR 22-JUL-1999; 990S-0145087.
PR 22-JUL-1999; 990S-0145089.
PR 22-JUL-1999; 990S-0145192.
PR 23-JUL-1999; 990S-0145145.
PR 23-JUL-1999; 990S-0145218.
PR 26-JUL-1999; 990S-0145224.
PR 27-JUL-1999; 990S-0145276.
PR 27-JUL-1999; 990S-0145913.
PR 27-JUL-1999; 990S-0145918.
PR 27-JUL-1999; 990S-0145919.
PR 28-JUL-1999; 990S-0145951.
PR 02-AUG-1999; 990S-0146386.
PR 02-AUG-1999; 990S-0146388.
PR 03-AUG-1999; 990S-0147038.
PR 04-AUG-1999; 990S-0147204.
PR 04-AUG-1999; 990S-0147302.
PR 05-AUG-1999; 990S-0147192.
PR 05-AUG-1999; 990S-0147260.
PR 06-AUG-1999; 990S-0147303.
PR 06-AUG-1999; 990S-0147416.
PR 09-AUG-1999; 990S-0147493.
PR 09-AUG-1999; 990S-0147935.
PR 10-AUG-1999; 990S-0148171.
PR 11-AUG-1999; 990S-0148319.
PR 12-AUG-1999; 990S-0148341.
PR 13-AUG-1999; 990S-0148565.
PR 13-AUG-1999; 990S-0148584.
PR 16-AUG-1999; 990S-0149368.
PR 17-AUG-1999; 990S-0149175.
PR 18-AUG-1999; 990S-0149426.
PR 20-AUG-1999; 990S-0149722.
PR 20-AUG-1999; 990S-0149723.
PR 20-AUG-1999; 990S-0149829.
PR 23-AUG-1999; 990S-0149902.
PR 23-AUG-1999; 990S-0149930.
PR 25-AUG-1999; 990S-0150566.
PR 26-AUG-1999; 990S-0150884.
PR 27-AUG-1999; 990S-0151065.
PR 27-AUG-1999; 990S-0151066.
PR 27-AUG-1999; 990S-0151080.
PR 30-AUG-1999; 990S-0151303.
PR 31-AUG-1999; 990S-0151438.
PR 01-SEP-1999; 990S-0151930.
PR 07-SEP-1999; 990S-0152363.
PR 10-SEP-1999; 990S-0153070.
PR 13-SEP-1999; 990S-0153758.
PR 15-SEP-1999; 990S-0154018.
PR 16-SEP-1999; 990S-0154039.
PR 20-SEP-1999; 990S-0154779.
PR 22-SEP-1999; 990S-0155139.
PR 23-SEP-1999; 990S-0155486.
PR 24-SEP-1999; 990S-0155659.
PR 28-SEP-1999; 990S-0156458.
PR 29-SEP-1999; 990S-0156596.
PR 04-OCT-1999; 990S-0157117.
PR 05-OCT-1999; 990S-0157753.
PR 06-OCT-1999; 990S-0157865.
PR 07-OCT-1999; 990S-0158029.
PR 08-OCT-1999; 990S-0158232.
PR 12-OCT-1999; 990S-0158369.
PR 13-OCT-1999; 990S-0159293.
PR 13-OCT-1999; 990S-0159294.
PR 13-OCT-1999; 990S-0159295.
PR 14-OCT-1999; 990S-0159329.
PR 14-OCT-1999; 990S-0159330.
PR 14-OCT-1999; 990S-0159331.

PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 23-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 5.98; Score 97.5; DB 21; Length 496;
Best Local Similarity 22.14; Pred. No. 0.027;
Matches 44; Conservative 30; Mismatches 86; Indels 39; Gaps 7;

OY 70 ISFDALFLGLIIRGYAALPNPQFSQFVKD-----LTKCYAMLLTYCGFVLDLVL 122
Db 312 ISFDALFLGLIIRGYAALPNPQFSQFVKD-----LTKCYAMLLTYCGFVLDLVL 365
OY 123 YHLNPKSSILEDYHEFLNINISFRNFIFAPITEEIFYTSMLTYLNL- PHSQLSYQO 181
Db 366 -----gsdftemyfvtfsfwnkyvyygfmllvftllyivtvcvtlygtyfllnaen 418
OY 182 LFWQPSLFFGLAHNAHAYEOLQEGSMVTVSILTTTCFQILVYTLFGL- TKFVFEVTRGG 239
Db 419 yhwqtsffsaa-----stavyvylsylyyyvktkmsgffqtsfyf---gy 462
OY 240 NLWCIIILHNLINMGFRG 258
Db 463 tmmfcldgldcgavgyly 481

Search completed: February 12, 2002, 09:27:26
Job time: 112 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 12, 2002, 09:32:01 ; Search time 43.94 Seconds
(without alignments)
1048.607 Million cell updates/sec

Title: US-09-165-460a-4
Perfect score: 1651
Sequence: 1 MGFSTFLVLYISISVLP.....ISLKDTQLTGVGTGPRITLL 315

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues
Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: SPREMEL_17:*
2: sp_archaea:*
3: sp_bacteria:*
4: sp_fungi:*
5: sp_human:*
6: sp_invertebrate:*
7: sp_mammal:*
8: sp_mhc:*
9: sp_organelle:*
10: sp_phage:*
11: sp_plant:*
12: sp_rodent:*
13: sp_virus:*
14: sp_vertebrate:*
15: sp_unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-----------|-----------------------|
| 1 | 227.5 | 13.8 | 314 | 11 Q9CSF8 | Q9CSF8 mus musculus |
| 2 | 207 | 12.5 | 225 | 11 Q9GK03 | Q9GK03 mus musculus |
| 3 | 113.5 | 6.9 | 339 | 5 Q18364 | Q18364 caenorhabditis |
| 4 | 111.5 | 6.8 | 225 | 8 Q9XN08 | Q9XN08 ceratitidis |
| 5 | 109 | 6.6 | 1363 | 5 Q10031 | Q10031 caenorhabditis |
| 6 | 104.5 | 6.3 | 235 | 5 Q9PKE0 | Q9PKE0 chlamydia m |
| 7 | 102.5 | 6.2 | 137 | 5 Q9NM17 | Q9NM17 leishmania |
| 8 | 102.5 | 6.2 | 1471 | 3 Q74805 | Q74805 schistosom |
| 9 | 99.5 | 6.0 | 224 | 8 Q9XMI3 | Q9XMI3 penaeus not |
| 10 | 98 | 5.9 | 256 | 2 Q84256 | Q84256 chlamydia t |
| 11 | 97.5 | 5.9 | 289 | 5 O17359 | O17359 caenorhabditis |
| 12 | 97.5 | 5.9 | 315 | 11 Q9EPB8 | Q9EPB8 mus musculus |
| 13 | 97.5 | 5.9 | 351 | 5 Q16945 | Q16945 alysia cal |
| 14 | 97.5 | 5.9 | 433 | 12 Q84483 | Q84483 paramyosin |
| 15 | 97.5 | 5.9 | 589 | 10 Q04091 | Q04091 arabidopsis |
| 16 | 97 | 5.9 | 483 | 2 Q9PBP9 | Q9PBP9 campylobact |
| 17 | 96.5 | 5.8 | 224 | 8 Q9MGNO | Q9MGNO drosophila |
| 18 | 96 | 5.8 | 328 | 2 Q9CEY3 | Q9CEY3 lactococcus |
| 19 | 96 | 5.8 | 333 | 4 Q9H2L2 | Q9H2L2 homo sapien |

| | | | | | |
|----|------|-----|-----|-----------|-----------------------|
| 20 | 96 | 5.8 | 346 | 8 Q9TCL5 | Q9TCL5 gymnostinop |
| 21 | 96 | 5.8 | 346 | 8 Q9G8K3 | Q9G8K3 psarocollus |
| 22 | 95.5 | 5.8 | 224 | 8 Q9MD63 | Q9MD63 drosophila |
| 23 | 95.5 | 5.8 | 434 | 11 Q54844 | Q54844 mus musculus |
| 24 | 95 | 5.8 | 274 | 2 Q9KFR4 | Q9KFR4 bacillus ha |
| 25 | 95 | 5.8 | 358 | 8 Q9GHR5 | Q9GHR5 penus bold |
| 26 | 95 | 5.7 | 500 | 5 Q22904 | Q22904 caenorhabditis |
| 27 | 94.5 | 5.7 | 224 | 8 Q9MG11 | Q9MG11 drosophila |
| 28 | 94.5 | 5.7 | 244 | 2 Q05525 | Q05525 bacillus su |
| 29 | 94.5 | 5.7 | 269 | 5 Q62262 | Q62262 caenorhabditis |
| 30 | 94.5 | 5.7 | 386 | 3 Q93875 | Q93875 candida alb |
| 31 | 94 | 5.7 | 346 | 5 Q16932 | Q16932 caenorhabditis |
| 32 | 94 | 5.7 | 347 | 2 Q9ZDJ0 | Q9ZDJ0 rickettsia |
| 33 | 94 | 5.7 | 432 | 10 Q24406 | Q24406 ameba |
| 34 | 94 | 5.7 | 554 | 5 Q9BPP5 | Q9BPP5 bombyx mand |
| 35 | 94 | 5.7 | 769 | 3 Q12244 | Q12244 saccharomyc |
| 36 | 93.5 | 5.7 | 224 | 8 Q9MGK7 | Q9MGK7 drosophila |
| 37 | 93 | 5.6 | 645 | 8 Q63620 | Q63620 balanogloss |
| 38 | 92.5 | 5.6 | 224 | 8 Q9MJC8 | Q9MJC8 drosophila |
| 39 | 92.5 | 5.6 | 224 | 8 Q9MGN5 | Q9MGN5 drosophila |
| 40 | 92.5 | 5.6 | 411 | 4 Q9Y2E8 | Q9Y2E8 homo sapien |
| 41 | 92.5 | 5.6 | 437 | 4 Q9Y507 | Q9Y507 homo sapien |
| 42 | 92.5 | 5.6 | 670 | 8 Q9T447 | Q9T447 dictyostell |
| 43 | 91.5 | 5.5 | 449 | 4 Q43934 | Q43934 homo sapien |
| 44 | 91.5 | 5.5 | 449 | 4 Q9NX15 | Q9NX15 homo sapien |
| 45 | 91.5 | 5.5 | 457 | 8 Q47876 | Q47876 alligator m |

ALIGNMENTS

RESULT 1
ID Q9CSF8 PRELIMINARY; PRT; 314 AA.
AC Q9CSF8;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE RAS AND A-FACTOR-CONVERTING ENZYME 1 HOMOLOG (S. CEREVISIAE)
DE (FRAGMENT).
GN RCE1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=EMBRYO;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Kono H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamada I.,
RA Saito T., Okazaki Y., Gojibori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Krehl P., Lewis S., Matsuo Y., Nikaido I., Paoletti G., Quackenbush J.,
RA Schriml L.M., Struhl F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barish G.,
RA Blake J., Boffelli D., Bojunga N., Carlini P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hune D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mezzarrelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Welter C., Whitaker C., Wilming L.,
RA Wyshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohsaki S.,
RA Hayashizaki Y.,
RT "Functional annotation of a full-length mouse cDNA collection."
RL Nature 409:685-690(2001).
DR EMBL: AK012946; BAB28566.1; -
DR MCD; MG1:1336895; Rcel.
FT NON_TER 1
SQ SEQUENCE 314 AA; 34390 MW; DE2624BB17E234B5 CRC64;

Query Match 13.8% Score 227.5; DB 11; Length 314;
 Best Local Similarity 30.2%; Pred. No. 1.6e-11;
 Matches 84; Conservative 41; Mismatches 82; Indels 71; Gaps 16;

QY 4 FSTFLVLLYISIVPLVATSOPEGRNDPRITKSMOKLTIMLSNLF-LVPELOSQ 62
 DB 25 FSCF-----SLACSYVGSILYVKSE--LPKDHPAVYKRR--STSVLVSSLSPLCVLWRE 76
 QY 63 LSSF---TSHISFPAFLGILIPGYAALPNPQSFQVVDLTKCVAMLT--LYCGPV 117
 DB 77 LTGIGPGTSLTLM-GRLEGIIPA--ALLP-----LLTMTILFLGPL 116
 QY 118 LDFV-----LYHLNPKSSILEDYHEFLINMS-----FRNFIFAPITEIF 159
 DB 117 MOLSMDCPCDLTDGLKAVLARS-----NARCLTDMRWLRNOYIAPTELV 163
 QY 160 YTSMLTLYLNLIPHSOLSYOQLFWQPSLFGLANHAYEOLQSGMTTYSILLTTCFQ 219
 DB 164 FRACMLPW---LAPCTGLG-PAVFTCP-LFEGVAHFHIIIEQLRPROSSVGSIFVSAFQ 218
 QY 220 ILTYTLFGGLTKFVFRVGGMLKCCIIILHALCINMGFP 257
 DB 219 FSYTAVFGATYATFLFIRT-GHLIGPVLCCHSFCNYMGFP 255

RESULT 2
 Q99K03 PRELIMINARY; PRT; 225 AA.
 AC Q99K03;
 DT 01-JUN-2001 (Tremblrel. 17, Created)
 DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
 DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
 DE UNKNOWN (PROTEIN FOR MCC:8174).
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 CX NCBI_Taxid=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Strausberg R.;
 RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: BC004060; AA04060.1;
 SO SEQUENCE 225 AA; 24791 MW; EFDDC2B6CE205B45 CRC64;

Query Match 12.5% Score 207; DB 11; Length 225;
 Best Local Similarity 31.8%; Pred. No. 6e-10;
 Matches 63; Conservative 27; Mismatches 50; Indels 58; Gaps 10;

QY 80 GIITGYAALPNPQSFQVVDLTKCVAMLT--LYCGPVLDV-----LYHLN 127
 DB 7 GIFPA--ALP-----LLTMTILFLGPIQLSMDPCDLTDGLKAVLA 47
 QY 128 PKSSILEDYHEFLINMS-----FRNFIFAPITEIFYSMLTLYLNLIPHSOLSY 179
 DB 48 PRS-----NARCLTDMRWLRNOYIAPTELVRAACMLPW---LAPCTGLG- 90
 QY 180 QQLFWQPSLFGLANHAYEOLQSGMTTYSILLTTCFOILTYTLFGGLTKFVFRVGTG 239
 DB 91 PAVFTCP-LFEGVAHFHIIIEQLRPROSSVGSIFVSAFQSYTAVFGATYATFLFIRT-G 148
 QY 240 NLWCCIIILHALCINMGFP 257
 DB 149 HLGIPVLCCHSFCNYMGFP 255

RESULT 3
 Q18364 PRELIMINARY; PRT; 339 AA.
 AC Q18364;
 DT 01-NOV-1996 (Tremblrel. 01, Created)

DT 01-JAN-1998 (Tremblrel. 05, Last sequence update)
 DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
 DE C33A12.10 PROTEIN.
 CN C33A12.10.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabdilitida; Rhabdilitoidea;
 OC Rhabdilitidae; Pelodierinae; Caenorhabditis.
 CX NCBI_Taxid=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Wilkinson J.;
 RL Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-94150718; PubMed-7906398;
 RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
 RA Briffeld J., Burton J., Connell M., Copey T., Cooper J., Coulson A.,
 RA Craxton M., Dear S., Du Z., Durbin R., Faveilo A., Fulton L.,
 RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
 RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
 RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
 RA Parsons J., Percy C., Rifken L., Koopra A., Saunders D., Showkeen R.,
 RA Smalton N., Smith A., Sonhammer E., Staden R., Sulston J.,
 RA Thierly-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
 RA Watson A., Weinstock L., Wilkinson-Spratt J., Wohlman P.,
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
 RT elegans".
 RL Nature 368:32-38(1994).
 DR EMBL: Z68493; CA92793.1;
 DR InterPro: IPR003839; DUF215.
 DR Pfam: PF02688; DUF215; 1.
 SO SEQUENCE 339 AA; 39177 MW; 05B7B6472ADC7DD5 CRC64;

Query Match 6.9% Score 113.5; DB 5; Length 339;
 Best Local Similarity 22.3%; Pred. No. 0.07;
 Matches 70; Conservative 41; Mismatches 102; Indels 101; Gaps 17;

QY 3 QFSFVLV-----YISIVPLVATRS--QPEGRKNDPRITKSMOKLTIMLS 51
 DB 28 QFSTFLAVPWIVIPSEFVICYLFKILYLSNMKSEGRKQHVFLVIS-LSQFTCL- 84
 QY 52 NLFLVPLFLOSLSSTTSHISF-----KDAFLGLGILPGY--YALPMPW----- 93
 DB 85 -LFLTDLMLPLPSTGTGFTSCASIPRNHYIKIIVLSALYRNILAMSPLLPVIRLIV 143
 QY 94 -----QFSQFVKDLTKCVAMLTLYCGVLDVLYHLNPKSSILEDYHEFLN 142
 DB 144 ICPKSHSKININIVKFAIPILCPFLCTEYFIPALG-VCKQFATP-----YFGS 193
 QY 143 IWSFRNFIFAPITEIFYSM--LTTLYLNLIPHSOLSYOQLFWQPS-----LFGGL 192
 DB 194 VW-----IYIKSAFGLRNSFFL-----YIVPMIASIITANGLFEKV 233
 QY 193 AHAHAYEOLQSGS--MTTYSILLTTCFOIL-YTLFGGLTKFV-----VR 236
 DB 234 AQAARAKLISAGSGSKTHISITMTLAMIIFYLTNCAFLVYFSGTYSIYSEIIR 293
 QY 237 TGNL-LWCCIIILHA 249
 DB 294 SFGNDLQFCVYVTA 307

RESULT 4
 Q9XN8 PRELIMINARY; PRT; 225 AA.
 AC Q9XN8;
 DT 01-NOV-1999 (Tremblrel. 12, Created)
 DT 01-NOV-1999 (Tremblrel. 12, Last sequence update)
 DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
 DE ATP SYNTHASE A CHAIN (EC 3.6.1.34).
 GN ATPase 6.
 OS Ceratitis capitata (Mediterranean fruit fly).

OG Mitochondrion.
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Tephritoidea; Tephritidae; Ceratitis.
 NC NCBL_Taxid=7213;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Spanos L., Koutroumbas G., Kotsyakis M., Louis C.;
 RT "The mitochondrial genome of the Mediterranean fruitfly, *Ceratitis capitata*,"
 RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: KEY COMPONENT OF THE PROTON CHANNEL; IT MAY PLAY A
 CC DIRECT ROLE IN THE TRANSLLOCATION OF PROTONS ACROSS THE MEMBRANE
 CC (BY SIMILARITY).
 CC -1- SUBUNIT: F-TYPE ATPASES HAVE 2 COMPONENTS. CF(1) - THE CATALYTIC
 CC CORE - AND CF(0) - THE MEMBRANE PROTON CHANNEL. CF(1) HAS FIVE
 CC SUBUNITS: ALPHA(3), BETA(3), GAMMA(1), DELTA(1), EPSILON(1). CF(0)
 CC HAS THREE MAIN SUBUNITS: A, B AND C (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
 CC -1- SIMILARITY: TO THE ATPASE A CHAIN FAMILY.
 DR EMBL: AJ242872; CAB45092.1; -
 DR InterPro: IPR000568; ATP_synth_A.
 DR Pfam: PF00119; ATP_synth_A.1.
 DR PRINTS: PR00123; ATPASEA.
 DR PROSITE: PS00449; ATPASE_A.1.
 DR CF(0); Hydrogen ion transport; Hydrolase; Mitochondrion;
 KM Transmembrane.
 SQ SEQUENCE 225 AA; 25113 MW; BC6D91F3AF3A915 CRC64;

Query Match 6.8%; Score 111.5; DB 8; Length 225;
 Best Local Similarity 22.0%; Pred. No. 0.067;
 Matches 58; Conservative 40; Mismatches 97; Indels 69; Gaps 12;

OY 49 LISNLFVPLFQISQSTSTSHISFKDAFLGIGIIPGYAALPNPWOQSFYKDLTKCYAM 108
 DB 1 MATNLFVFDPSSTIFNLS--LNMSTFLGILLIPSAWMLPSSV----- 43
 OY 109 LLLTYCGVLDVFLYHLNPKSSILEDYHEFLN-----WSFNFIAPITTEIFYSM 163
 DB 44 -----HIF--WNSILMTLHKEFKTLLEGSGHSGSTFIVSL-----FSLI 81
 OY 164 LLLTYLNLIPHSQSLYQOL-----FMQPSLFFG-LAHANHAVEQLQSGSMITVSIL 213
 DB 82 LFNFMGLFPIYITSTSLTLTLTLALPLWLCFMYLGMVINTQNHFTLVYQGPRAVLMP 141
 OY 214 LITCGQIITLFGGLTKFVYVTRGNIWCCIIILHALCINMGFPPSKLNLFYVD-KK 272
 DB 142 FMVCIETISNIRPGT--LAVRLAAM--IAGHLTLTLTGNTGPS--LSFMIVSILL 192
 OY 273 AGRISKL-----VSIMNKCYPALL 291
 DB 193 IGQIALLVESAVALIQSYVAVL 216

RESULT 5
 O10031 PRELIMINARY; PRT; 1363 AA.
 AC Q10031;
 DT 01-NOV-1996 (TREMBlrel. 01, Created)
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
 DE PUTATIVE SERINE/THREONINE-PROTEIN KINASE C27D6.2 IN CHROMOSOME II
 DE (NC 2.7.1.-).
 GN C27D6.2.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 OX NCBL_Taxid=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RA Ding H.;

RL Submitted (JUL-1995) to the EMBL/GenBank/DBJ databases.
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
 CC -1- SIMILARITY: TO THE SER/THR FAMILY OF PROTEIN KINASES.
 DR EMBL: U23179; AAC46722.1; -
 DR WormPep: C27D6.2; CE01812.
 DR InterPro: IPR002106; AA_tRNA_ligase-II.
 DR InterPro: IPR000719; Euk_kinase.
 DR InterPro: IPR002290; Ser_thr_kin_acctsite.
 DR InterPro: IPR002184; Srb.
 DR Pfam: PF00069; Pkinase.1.
 DR Pfam: PF02175; Srb.5.
 DR SMART: SM00220; S_TKC.1.
 DR PROSITE: PS00339; AA_TRNA_LIGASE_II_2; UNKNOWN_4.
 DR PROSITE: PS00107; PROTEIN KINASE ATP; UNKNOWN_1.
 DR PROSITE: PSS0011; PROTEIN KINASE DOM.1.
 DR PROSITE: PS00108; PROTEIN_KINASE_ST.1.
 DR ATP-binding; Hypothetical protein; Serine/threonine-protein kinase;
 KM transferase; Transmembrane.
 FT TRANSMEM 22 42 POTENTIAL.
 FT TRANSMEM 66 86 POTENTIAL.
 FT TRANSMEM 141 161 POTENTIAL.
 FT TRANSMEM 188 208 POTENTIAL.
 FT TRANSMEM 279 299 POTENTIAL.
 FT TRANSMEM 380 400 POTENTIAL.
 FT TRANSMEM 415 435 POTENTIAL.
 FT TRANSMEM 460 480 POTENTIAL.
 FT TRANSMEM 498 518 POTENTIAL.
 FT TRANSMEM 544 564 POTENTIAL.
 FT TRANSMEM 592 612 POTENTIAL.
 FT TRANSMEM 635 655 POTENTIAL.
 FT TRANSMEM 671 691 POTENTIAL.
 FT DOMAIN 679 682 POLY-LEU.
 FT TRANSMEM 716 736 POTENTIAL.
 FT TRANSMEM 754 774 POTENTIAL.
 FT TRANSMEM 832 852 POTENTIAL.
 FT TRANSMEM 867 887 POTENTIAL.
 FT TRANSMEM 913 933 POTENTIAL.
 FT TRANSMEM 952 972 POTENTIAL.
 FT TRANSMEM 999 1019 POTENTIAL.
 FT DOMAIN 1068 1363 POTENTIAL.
 FT NP_BIND 1074 1082 ATP (BY SIMILARITY).
 FT ACT_SITE 1193 1193 BY SIMILARITY.
 FT BINDING 1097 1097 ATP (BY SIMILARITY).
 SQ SEQUENCE 1363 AA; 157850 MW; 7A030E80C96C0A6 CRC64;

Query Match 6.6%; Score 109; DB 5; Length 1363;
 Best Local Similarity 21.5%; Pred. No. 0.73;
 Matches 67; Conservative 46; Mismatches 110; Indels 88; Gaps 16;

OY 18 VLPPIATSQPGSKRDNPRITKSMQKITMLISNLPVPLQSLSTSHISFKDAFL 77
 DB 219 VAPLFITSMCD-----LITRPSLYK-----VGNLSLTLEMTQM----- 252
 OY 78 GLGITIPGYAALPNPWOQSFVK-DLTKCYAMLLTYCGVLDV-----VLVHLNLP 128
 DB 253 -----IMPLQFSIERIALSMKTSYENVRT-FLGPIVLVTLTGIDALLHYVFRD 301
 OY 129 KSSILEDYHEFLNIMSEFNIFAPITE-----EIFYSM-LTTYLNLIPHSQ 177
 DB 302 EK-FED-----SPISFALVPENSAIPFNSYFELLVAEIGNFCIFLIVHSKF 350
 OY 178 S---YQQLFWQPS---LFGLAHHAHVEQLQSGSMITVSILLTTCQIILTYTTFGLTK 231
 DB 351 KARVHOMLENDVSCVCELAQOLAY-HPVYRSSQFMSMLVSSISIPALYFIRKIF----- 404
 OY 232 FVFPRTGNIWCCIIILHALCNI-----MGFPPSKLNLFYVDKKAGRISKVSIWNK- 285
 DB 405 --FLHFGNKLCLLIYFICHLFSLMALCFAPFQIPLFPVTSKQCLLIWTLTFKNGQI 462
 OY 286 CYFALLVGLIT 296
 DB 463 CSFLLTSSML 473


```

Db 579 FISLONSKNEFIYLLDFY--MOLVSSONKNPKRTAISRRKPTLSMKSSSL---SOLMT 633
OY 62 QLSSTSHISFKDAFLGIIIPGYAAL-PN-----PMQFSQPVKDLTICVAMLLTY--C 114
Db 634 TVSSSTNVA-----YIRCIKPNKEKLPMTESP-----DPAVLSQLRAC 669
OY 115 GPVLDVFLVYHLNPKSSILEDYHEFLNMFNFIAPITEEIEFTYSMLTTLTNLIPH 174
Db 670 GVFETIRISSIGFAPRSEYEFAHRFRLLSSKEM-----EDDNKKTLTNVNSVIPH 722
OY 175 SOLSYQ---OLFQPSLFEFLAHNAHNA-----YEQLQEGSMTTVSIILTTTCFOI-L 221
Db 723 DNLNFQVRSKRIFFRSNIIGNFEAHNRATCSKSVTLQSAIRGFEFTREKQRTVYKFIKL 782
OY 222 YTLFEGGLTKFVFYR 236
Db 783 QSVIMGWLTTRKPER 797

RESULT 9
O9XMT3 PRELIMINARY; PRT: 224 AA.
ID O9XMT3;
AC O9XMT3;
DT 01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE ATP SYNTHASE A CHAIN (EC 3.6.1.34).
GN ATP6.
OS Pennaeus notialis.
OC Mitochondrion.
OC Eukaryota; Metazoa; Arthropoda; Crustacea; Malacostraca;
OC Eumalacostraca; Eucarida; Decapoda; Dendrobranchiata; Penaeoidea;
OC Penaeidae; Pennaeus.
OX NCBI_TaxID=44108;
RN [1]
RP SEQUENCE FROM N.A.
RA Garcia-Machado E., Denneboud N., Oliva-Suarez M., Mounolou J.C.,
RA Mounolou J.C., Monnerot M.;
RT "Partial sequence of the shrimp Pennaeus notialis mitochondrial
RT genome.";
RL C. R. Acad. Sci., D. Sci. Nat. 319:473-486(1996).
RN [2]
RP SEQUENCE FROM N.A.
RA Garcia-Machado E.;
RL Thesis (1997), Paris-Sud, UFR d'Orsay, France.
RN [3]
RP SEQUENCE FROM N.A.
RA Garcia-Machado E., Pempera M., Denneboud N., Oliva-Suarez M.,
RA Mounolou J.C., Monnerot M.;
RT "Mitochondrial genes collectively suggest the paraphyly of Crustacea
RT with respect to Insecta.";
RL Submitted (MAR-1999) to the EMBL/Genbank/DBJ databases.
CC -1- FUNCTION: KEY COMPONENT OF THE PROTON CHANNEL. IT MAY PLAY A
CC DIRECT ROLE IN THE TRANSLLOCATION OF PROTONS ACROSS THE MEMBRANE
CC (BY SIMILARITY).
CC -1- SUBUNIT: F-TYPE ATPASES HAVE 2 COMPONENTS. CF(1) - THE CATALYTIC
CC CORE - AND CF(0) - THE MEMBRANE PROTON CHANNEL. CF(1) HAS FIVE
CC SUBUNITS: ALPHA(3), BETA(3), GAMMA(1), DELTA(1), EPSILON(1). CF(0)
CC HAS THREE MAIN SUBUNITS: A, B AND C (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
CC -1- SIMILARITY: TO THE ATPASE A CHAIN FAMILY.
DR EMBL: X84350; CAB40367.1; -
DR InterPro: IPR000568; ATP_synth_A.
DR Pfam: PF00119; ATP_synth_A; 1.
DR PRINTS: PR00123; ATPASEA.
DR PROSITE: PS00449; ATPASE_A; 1.
KW CF(0): Hydrogen ion transport; Mitochondrion; Transmembrane.
SQ SEQUENCE 224 AA; 2474 MW; E903131029A715A5 CRC64;

Query Match 6.0%; Score 99.5; DB 8; Length 224;
Best Local Similarity 22.2%; Pred. No. 0.68;
Matches 57; Conservative 41; Mismatches 78; Indels 81; Gaps 15;

```

```

OY 51 SNLFLVPELQSLSTSHISFKDAFLGIIIPGYAALPNPWO--FSQFVKDLTKCYAM 108
Db 13 SIFMLP-----LWVSTFLGVMLPMLYAMPSSMLMLVATLTKKEKT 60
OY 109 LL-TLYCGPVLDV-LYHLNPKSSILEDYHEFLNMFNFIAPITEEIEFTYSMLLT 166
Db 61 LIGSSHFGTTLMPVSLFSLI-----VENNFLGLLP--YIFTS-----TSHLVM 101
OY 167 TYLNLIPIHSQSLSYQOLFQPSLFEFLAHNAHNAEQ---OEGSMTTVSIILTTTCFOI-L 221
Db 102 TLSIALP-----LWVAFMLFGWLNHTOHMFANLPOGTPGALMPFVLIETISNVI 152
OY 222 YTLFEGGLTKFVFYRGTGMLMCCILHALCNMGPPSRILNHTVYVDDKKAGRSKLVS 281
Db 153 RP---GTLA---VRLAANM---IAGHLTLTLGSGPS--LSATLI----- 187
OY 282 IWNKCYFALLVGLISL 298
Db 188 -----SMLIGQILL 197

RESULT 10
O84256 PRELIMINARY; PRT: 256 AA.
ID O84256;
AC O84256;
DT 01-NOV-1998 (TREMBLrel. 08, Created)
DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE HYPOTHETICAL 28.4 KDA PROTEIN.
GN CT254.
OS Chlamydia trachomatis.
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=813;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=D/UM-3/CX;
RX MEDLINE=9900809; PubMed=9784136;
RA Stephens R.S., Kalman S., Lammel C.J., Fan J., Marathe R., Aravind L.,
RA Mitchell W.P., Olinger L., Tatusov R.L., Zhao Q., Koonin E.V.,
RA Davis R.W.;
RT "Genome sequence of an obligate intracellular pathogen of humans:
RT Chlamydia trachomatis.";
RL Science 282:754-759(1998).
DR EMBL: AE001298; AAC67847.1; -
DR InterPro: IPR003675; AD1.
DR Pfam: PF02517; AD1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 256 AA; 28378 MW; 25BEB45DE4B53B83 CRC64;

Query Match 5.9%; Score 98; DB 2; Length 256;
Best Local Similarity 22.9%; Pred. No. 1;
Matches 60; Conservative 34; Mismatches 110; Indels 58; Gaps 9;
OY 1 MLOESTFLVLYISYVPLVATISOPGSKRDNRTIKSRQKLTIMLSLFLVPL- 59
Db 1 MWGLGLLVFVGISIFLVNGFVWPKPK-----HTRLPFLVLCGAVLLLPALL 51
OY 60 -QSLSSTSHISFQDAF-----LGL-----GIIPGYAALPNPQPS----- 96
Db 52 PTGDDLSISKVSFGICIVFSAFLGLGAPRPTOSILYAGRTSSTPMLAIAAGARM 111
OY 97 -QFVKDLTKCYAMLLTLYCGPVLDVLYHLNPKSSILEDYHEFLN---IWSFRNFIE 151
Db 112 WYITTAQIILARIQL---VLPILPKQLFQEVITEIEEISQSLSTRPDLIYICIAML 167
OY 152 APIITEIEFTYSMLTTLTN-----LIPIHSQSLSYQOLFQPSLFEFLAHNAH 196
Db 168 IPIAEIEFPGILQTFEENFNTSRSAVYTSLSIFAITHIASLGSFIVPILVPSLCA 227
OY 197 HAYEQLQEGSMTTVSIILTTTC 217

```

Db 228 FIVEKVRHIAAPVTLHLFNSC 249

RESULT 11

017359 PRELIMINARY; PRT; 289 AA.
 AC 017359;
 DT 01-JAN-1998 (Tremblrel. 05, Created)
 DT 01-JAN-1998 (Tremblrel. 05, Last sequence update)
 DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
 DE COSE4.5 PROTEIN.
 GN COSE4.5.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
 CC Rhabditidae; Peloderinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RX MEDLINE=94150718; PubMed=7906398;
 RA Wilson R., Almscough R., Anderson K., Baynes C., Berks M.,
 Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
 Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
 Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
 Jones M., Kershaw J., Kirsten J., Lalster N., Latreille P.,
 Lightning J., Lloyd C., Mcmurray A., Mortimore B., O'Callaghan M.,
 Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Shownkeen R.,
 Smidson N., Smith A., Sonhammer E., Staden R., Sulston J.,
 Thierly-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
 Watson A., Weinstock L., Wilkinson-Sproat J., Woldman P.,
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
 elegans."
 RT Nature 368:32-38(1994).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RA Blanchard M., Kramer J., Gibson A.;
 RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RA Waterston R.;
 RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF026209; AAB71274.1;
 DR InterPro: IPR000168; 7TM_nematode.
 DR InterPro: IPR003002; 7TM_chemorecept_1.
 DR Pfam: PF01461; 7tm_4; 1.
 SO SEQUENCE 289 AA; 32908 MW; FFC7953558200B31 CRC64;

Query Match 5.9%; Score 97.5; DB 5; Length 289;
 Best Local Similarity 23.4%; Pred. No. 1.3;
 Matches 60; Conservative 26; Mismatches 71; Indels 99; Gaps 13;

QY 95 FSQFVKDLTKCVAMLLLYCGPVLDFVLYHLINPKSSILEDYFHE---EPLINMSFRNFT 150
 Db 15 FSQFVY-----FILLYLLRAKAG-----KHLGPIGFKNLSEF-NFF 49
 QY 151 FAPITEEFY-----TSMELT-----TYNLILPHQSLSYQOLF----- 183
 Db 50 Y-----TFYLGIFGASGLCTSLCTEFMRFRYIYCOOKYHLLEGRKRLALLFLEPTTI 103
 QY 184 ---NQPISLFGSLAHNAHAYEOL---QEGSMTVSILLTTFOILYTLTLEGLTFKVFVR 236
 Db 104 SITWTCGCFGLITSEKQELARIPQENFGEDSNVLMFVSGQY-----WIRGA 152
 QY 237 TGNLMC-----CIILALCNIMGFPGPSRLNLHFTVVDKRAKGRISLVSIMNKCYPAL 290
 Db 153 NGEKIMCLRDGCFGLIGLIGCSVISFCGLKT---FKMIEVQSGMSQATALNKLQPLT 208
 QY 291 LVLGILSLKQTLQTLV 306
 Db 209 L-----TLQTL 215

RESULT 12

09EPB8 PRELIMINARY; PRT; 315 AA.
 AC 09EPB8;
 DT 01-MAR-2001 (Tremblrel. 16, Created)
 DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
 DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
 DE VN7 (VOMERONASAL RECEPTOR VIR9).
 GN VIR9.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Lane R.P., Rowen L., Friedman C., Trask B.J., Hood L.;
 RT "Genomic characterization of the murine chromosome 6D1 VNO olfactory
 receptor cluster."
 RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-129/SVJ;
 RX MEDLINE=20568485; PubMed=11116090;
 RA Del Punta K., Rothman A., Rodriguez I., Mombaerts P.;
 RT "Sequence diversity and genomic organization of Vomeronasal Receptor
 Genes in the Mouse."
 RL Genome Res. 10:1958-1967(2000).
 DR EMBL; AF129005; AAG3282.1;
 DR EMBL; AF291488; AAG42082.1;
 DR InterPro: IPR000276; GPCR_Rhodopsn.
 DR PROSITE: PS50262; G_PROTEIN_RECEP_FL2; 1.
 KW Receptor.
 SO SEQUENCE 315 AA; 35853 MW; 1D6A41D0FEBB6C93 CRC64;

Query Match 5.9%; Score 97.5; DB 11; Length 315;
 Best Local Similarity 20.4%; Pred. No. 1.4;
 Matches 69; Conservative 50; Mismatches 111; Indels 109; Gaps 15;

QY 5 STEPLVLYISISVPLVATSOPEGSKRDNPRTKSRMOKLTMILNISLVLVPLQSOIS 64
 Db 28 NSFLLFHI-FKYI-----RGQRSHRIDIPGILSLIHLVMIAMSLV 69
 QY 65 STSHISFKDAFLGILGILGYAALPNW-----QFOQFVADLTKCVAMLLTL 112
 Db 70 AT-----DIFW-----FWGRGDTCTCKCVISLYRFCRSLSLCATSLTS- 107
 QY 113 YCGPVLDFVLYHLINPKSSILEDYFHE-----FLINMSFRNFTFADITEIEFYS 162
 Db 108 ----ILQAV---TLNPRNSCLEKFRKSPHYMGLCLFLSV--FYTFSSPLATYITKS 158
 QY 163 ML-----LTYNLILPHQSLSYQOLF---WQPSLFGSLAHNAHAY-----EQL 202
 Db 159 NLTPSPFYITTSCLAPMSYFHLTWFIILTSRDVIFVGLMLSSGYMTFLGRHKKS 218
 QY 203 QEGSMTVS-----ILLTFOILYTLTLEGLTFKVFVRFGMLMCCIL- 247
 Db 219 QPLHTSFLKPSAEKRAMRTILCLMSFEVLMYR--LDSIVSYRSIDGQIFCVHIFT 276
 QY 248 -HALCNIMGFPGPSRLNLHFTVVDKRAKGRISLVSIMNKCYPAL 285
 Db 277 AHGYATVSPFLILSTEKYIINFRSTGRMVTITILLNR 315

RESULT 13
 Q16945 PRELIMINARY; PRT; 351 AA.
 AC 016945;
 DT 01-NOV-1996 (Tremblrel. 01, Created)
 DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
 DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)

| | | | | |
|-----------------------|------------------|----------------|------------|-------------|
| Query Match | 5.9%; | Score 97.5; | DB 5; | Length 351; |
| Best Local Similarity | 25.4%; | Pred. NO. 1.6; | | |
| Matches 61; | Conservative 30; | Mismatches 84; | Indels 65; | Gaps 13 |

| | | |
|--------------|--------|-----|
| RESULT | 14 | |
| 084483 | | |
| ID | 084483 | |
| 084483 | | |
| PRELIMINARY; | | |
| PRT; | 433 | AA. |

```
.query match      5.98;  score 97.5;  db 12;  length 433;
```

| | | |
|--------|--------|--------------|
| RESULT | 15 | |
| ID | 004091 | PRELIMINARY; |
| | | PRT; 589 AA. |

| | | | | | | | |
|-----------------------|-------|--------------|---------|------------|----|--------|-----|
| Query Match | 5.9% | Score | 97.5 | DB | 10 | Length | 589 |
| Best Local Similarity | 22.1% | Pred. | No. 2.8 | | | | |
| Matches | 44 | Conservative | 30 | Mismatches | 86 | Indels | 39 |
| | | | | | | Gaps | 7 |

QY 70 ISFKAFALGGLGIPYVALPMPQFSGFVK------LTFQVMALLLXGCPVLEFL 122
 Db 405 ISFLLALLGTVGRMWSGAPNPNCKVKTTPRPIPEKKWLTLSVYSLSM----GGLLPF--- 458
 QY 123 YHLNPKSSILEDYFHEFLNWSFSPNFIAPITTEIEFYTSMLLTYVANI-PSOLSXYOO 181
 Db 459 -----GSIFIEYFVFSTSPNRYKYVYYGFMLLVEFLVLTVCVTIVGTYPIYLNAMN 511
 QY 182 LFWQSLFFGGLAHNAHAYEQLOEGSMTTYVSIILTTQFOILYTTLFGGL--TFYFVYVRGG 239
 Db 512 YHMQTSFEFSA-----STAVYVYLYSIYYVYVYVTKKMSGFQTSFYF---GY 555
 QY 240 NLMCCIILHALCNI MGFRG 258
 Db 556 TMMELGLGILGAGVYLG 574

Tue Feb 12 16:22:39 2002

us-09-165-460a-4.rpt

Page 8

Search completed: February 12, 2002, 09:32:03
Job time: 350 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 12, 2002, 09:26:07 ; Search time 21.66 Seconds
(without alignments)
327.264 Million cell updates/sec

Title: US-09-165-460a-4

Perfect score: 1651
Sequence: 1 MLOFSTFELVLYISISYVLP.....ISLKDTLQTLVGPGRYRTL 315

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued_Patents_Aa: *
1: /cgn2_6/ptodata/2/1aa/5A-COMB.pep: *
2: /cgn2_6/ptodata/2/1aa/5B-COMB.pep: *
3: /cgn2_6/ptodata/2/1aa/6A-COMB.pep: *
4: /cgn2_6/ptodata/2/1aa/6B-COMB.pep: *
5: /cgn2_6/ptodata/2/1aa/PCTUS-COMB.pep: *
6: /cgn2_6/ptodata/2/1aa/Backfile1.pep: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|-------------------|--------------------|
| 1 | 228.5 | 13.8 | 329 | US-09-022-669-2 | Sequence 2, Appli |
| 2 | 227.5 | 13.8 | 329 | US-09-262-749-2 | Sequence 2, Appli |
| 3 | 96.5 | 5.8 | 456 | US-09-058-389A-2 | Sequence 2, Appli |
| 4 | 91.5 | 5.5 | 470 | US-09-292-071-25 | Sequence 25, Appli |
| 5 | 91.5 | 5.5 | 470 | US-09-292-069A-25 | Sequence 25, Appli |
| 6 | 91.5 | 5.5 | 471 | US-07-817-920-8 | Sequence 8, Appli |
| 7 | 91.5 | 5.5 | 471 | US-08-370-542-7 | Sequence 8, Appli |
| 8 | 91.5 | 5.5 | 471 | US-08-117-006-8 | Sequence 8, Appli |
| 9 | 91.5 | 5.5 | 471 | US-08-216-594-8 | Sequence 8, Appli |
| 10 | 91.5 | 5.5 | 471 | US-08-542-358-7 | Sequence 7, Appli |
| 11 | 91.5 | 5.5 | 471 | US-08-244-434-2 | Sequence 2, Appli |
| 12 | 91.5 | 5.5 | 471 | US-09-018-351-7 | Sequence 7, Appli |
| 13 | 91.5 | 5.5 | 471 | US-09-032-742-8 | Sequence 8, Appli |
| 14 | 91.5 | 5.5 | 471 | PCT-US93-00149-8 | Sequence 8, Appli |
| 15 | 91.5 | 5.5 | 471 | US-07-996-772A-11 | Sequence 11, Appli |
| 16 | 89.5 | 5.4 | 471 | US-09-032-742-2 | Sequence 2, Appli |
| 17 | 89 | 5.4 | 3174 | US-08-477-451-3 | Sequence 3, Appli |
| 18 | 87.5 | 5.3 | 445 | US-08-630-118A-2 | Sequence 2, Appli |
| 19 | 87.5 | 5.3 | 445 | US-08-838-399-2 | Sequence 2, Appli |
| 20 | 87.5 | 5.3 | 445 | US-09-235-839-2 | Sequence 2, Appli |
| 21 | 84 | 5.1 | 462 | US-08-477-451-24 | Sequence 24, Appli |
| 22 | 83.5 | 5.1 | 445 | US-08-630-118A-4 | Sequence 4, Appli |
| 23 | 83.5 | 5.1 | 445 | US-08-630-118A-6 | Sequence 6, Appli |
| 24 | 83.5 | 5.1 | 445 | US-08-838-399-4 | Sequence 4, Appli |
| 25 | 83.5 | 5.1 | 445 | US-08-838-399-6 | Sequence 6, Appli |
| 26 | 83.5 | 5.1 | 445 | US-09-003-199-2 | Sequence 2, Appli |
| 27 | 83.5 | 5.1 | 445 | US-09-003-199-21 | Sequence 21, Appli |

| | | | | | |
|----|------|-----|-----|-------------------|--------------------|
| 28 | 83.5 | 5.1 | 445 | US-09-003-199-23 | Sequence 23, Appli |
| 29 | 83.5 | 5.1 | 445 | US-09-235-839-4 | Sequence 4, Appli |
| 30 | 83.5 | 5.1 | 445 | US-09-235-839-6 | Sequence 6, Appli |
| 31 | 83.5 | 5.1 | 445 | US-09-040-958-2 | Sequence 2, Appli |
| 32 | 83.5 | 5.1 | 445 | US-09-040-958-4 | Sequence 4, Appli |
| 33 | 83.5 | 5.1 | 445 | US-08-349-025-4 | Sequence 4, Appli |
| 34 | 83.5 | 5.1 | 445 | US-08-566-096A-4 | Sequence 4, Appli |
| 35 | 83.5 | 5.1 | 445 | US-08-668-650B-4 | Sequence 4, Appli |
| 36 | 83.5 | 5.1 | 445 | US-08-349-025-2 | Sequence 2, Appli |
| 37 | 83.5 | 5.1 | 445 | US-08-566-096A-2 | Sequence 2, Appli |
| 38 | 83.5 | 5.1 | 445 | US-08-668-650B-2 | Sequence 2, Appli |
| 39 | 83.5 | 5.1 | 445 | PCT-US95-15646-2 | Sequence 2, Appli |
| 40 | 83.5 | 5.1 | 445 | PCT-US95-15646-4 | Sequence 4, Appli |
| 41 | 83 | 5.0 | 821 | US-09-422-869-24 | Sequence 24, Appli |
| 42 | 82.5 | 5.0 | 314 | US-08-988-876-7 | Sequence 7, Appli |
| 43 | 82 | 5.0 | 379 | US-08-118-270-32 | Sequence 32, Appli |
| 44 | 82 | 5.0 | 379 | PCT-US93-08528-32 | Sequence 32, Appli |
| 45 | 81.5 | 4.9 | 326 | US-09-058-389A-3 | Sequence 3, Appli |

ALIGNMENTS

RESULT 1
US-09-022-669-2
Sequence 2, Application US/09022669
Patent No. 610717
GENERAL INFORMATION:
APPLICANT: KIRBY, KRISTINE
APPLICANT: SOUTHAN, CHRISTOPHER
APPLICANT: KNAB, ANNE
TITLE OF INVENTION: Human RCE1
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESSES:
ADDRESS: RATNER & PRESTIA
STREET: P.O. BOX 980
CITY: VALLEY FORGE
STATE: PA
COUNTRY: USA
ZIP: 19482
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTED for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/022.669
FILING DATE: 12-FEB-1998
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 97304437.3
FILING DATE: JUNE 24, 1997
ATTORNEY/AGENT INFORMATION:
NAME: PRESTIA, PAUL F
REGISTRATION NUMBER: 23,031
REFERENCE/DOCKET NUMBER: GH-70379
TELECOMMUNICATION INFORMATION:
TELEPHONE: 610-407-0700
TELEFAX: 610-407-0701
TELEX: 846169
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 329 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-022-669-2
Query Match 13.8%; Score 228.5; DB 3; Length 329;
Best Local Similarity 30.6%; Pred. No. 6.4e-17;
Matches 83; Conservative 39; Mismatches 82; Indels 67; Gaps 15;

Page 2

| | | |
|----|----|--------------------|
| QY | 11 | LYISYVLPYATSQPEGSK |
| | 1 | 1 |

US-09-058-389A-2

```

0Y      11  LYSISVLPVLTATSOEGSKRDNPRITKSRMÖKLTIMLISNF-LVPPLOSLSSTTSH 69
      |::|||  |  ||:|  ||  |  ::::|:|  |  3:|:

```

RESULT 4
 US-09-292-071-25
 : Sequence 25, Application US/09292071
 : Patent No. 6107324
 : GENERAL INFORMATION:
 : APPLICANT: Behan, Dominic
 : APPLICANT: Chalmers, Derrick
 : TITLE OF INVENTION: No. 6107324-Endogenous, Constitutively Activated
 : TITLE OF INVENTION: Human Serotonin Receptors and Small Molecule Modulators Thereof
 : NUMBER OF SEQUENCES: 33
 : CORRESPONDENCE ADDRESS:
 : ADDRESSEE: Arena Pharmaceuticals, Inc.
 : STREET: 6166 Nancy Ridge Drive
 : CITY: San Diego
 : STATE: CA
 : COUNTRY: USA
 : ZIP: 92121
 :
 : COMPUTER READABLE FORM:
 : MEDIUM TYPE: Floppy disk
 : COMPUTER: IBM PC compatible
 : OPERATING SYSTEM: PC-DOS/MS-DOS
 : SOFTWARE: PatentIn Release #1.0, Version #1.30
 : CURRENT APPLICATION DATA:
 : APPLICATION NUMBER: US/09/292,071
 : FILING DATE: April 14, 1999
 : CLASSIFICATION:
 : ATTORNEY/AGENT INFORMATION:
 : NAME: Mark J. Rosen
 : REGISTRATION NUMBER: 39,822
 : TELECOMMUNICATION INFORMATION:
 : TELEPHONE: (215) 564-6525
 : TELEFAX: (215) 568-3439
 : INFORMATION FOR SEQ ID NO: 25:
 : SEQUENCE CHARACTERISTICS:
 : LENGTH: 470 amino acids
 : TYPE: amino acid
 : STRANDEDNESS:
 : TOPOLOGY: not relevant
 : MOLECULE TYPE: protein
 :
 : US-09-292-071-25

```

1      RESULT      5
2      US-09-292-069A-25
3      Sequence 25, Application US/09292069A
4      Patent No. 6140509
5      GENERAL INFORMATION:
6      APPLICANT: Behan, Dominic P
7      APPLICANT: Chalmers, Derek T
8      APPLICANT: Foster, Richard J
9      APPLICANT: Glen, Robert C
10     APPLICANT: Lawless, Michael S
11     APPLICANT: Liu, Qian W
12     APPLICANT: Russo, Joseph F
13     APPLICANT: Smith, Julian R
14     APPLICANT: Thomsen, William J
15     TITLE OF INVENTION: No. 6140509-Endogenous, Constitutively Activated Human
16     TITLE OF INVENTION: Serotonin Receptors And Small Molecule Modulators
17     TITLE OF INVENTION: Thereof
18     FILE REFERENCE: AREN0033
19     CURRENT APPLICATION NUMBER: US/09/292,069A
20     CURRENT FILING DATE: 1999-04-14
21     PRIOR APPLICATION NUMBER: 60/090,783
22     PRIOR FILING DATE: 1998-06-26
23     PRIOR APPLICATION NUMBER: 60/112,909
24     PRIOR FILING DATE: 1998-12-18
25     PRIOR APPLICATION NUMBER: 60/123,000
26     PRIOR FILING DATE: 1999-03-05
27     NUMBER OF SEQ ID NOS: 33
28     SOFTWARE: PatentIn Ver. 2.1
29     SEQ ID NO 25
30     LENGTH: 470
31     TYPE: PRT
32     ORGANISM: Artificial Sequence
33     FEATURE:
34     OTHER INFORMATION: Description of Artificial Sequence: No. 6140509e1
35     US-09-292-069A-25

```

```
Query Match      91.5%   Score 91.5; DB 4    length 470;  
Best Local Similarity     20.4%; Pred. No. 0.1;  
Matches        69; Conservative       52; Mismatches    77; Indels    141; Gaps    21
```

OY 45 LTIMISMLVLPFQSQSLSTTSH-----ISKDAFLGLGITP-----GY-----85
|||.:|||.:.|.|.:|||.:.|.|.:|||.:.|.|.:|||.:.|.|.
Db 93 LYIIMAVS-----LEKRLQNATNIFLMSLAADMLGTLVPVSMTLLTYLRWPLPSK 145
|||||.....

OY 86 -----YAALPNPWQSOF---VKDLTKCVAMLLTTYCG 115
|..|...|.

Db 146 LCAWIVLYDLVLFSTASIMHLCALISDRVYALONPIHHSRFSNRTKAFKIIA-VWTISVG 204
Qy 116 PVLDVLYHLNPK-----SILEDYHEFLINTSFNRFIAPRTTEIFTSMLLTYL 169
Db 205 ISMPIPVGLDDSKVKEGSCILAD--DNEVLIGSFVSF-FIPL-----TIMVITYFL 255
Qy 170 NL-----IPHSOLSYOOLFW-----QPSLFGLAHAHAY 199
Db 256 TIKSLQKATLCVSDLGTRAKLASFLPQSSLSSEKLFORSIHREPGSYTG----- 307
Qy 200 EQLQSGMTVYSILTTGFOILYTTLFGGLTKFVFRGTGNLMC-----CITLHALC 251
Db 308 ----RRTMOSISNEQKAC-KYL-----GIVFELFV-----VWMCFFITINIMAVICKRSC 352
Qy 252 N--IMGPPGSRNLHFTVVDKKGRIKLY-SIWNKCY 287
Db 353 NEDVIG----ALLNV-FWIGYLSAVNPPLYTYTFENKTY 386

RESULT 6

US-07-817-920-8
Sequence 8, Application US/07817920
Patent No. 5360735

GENERAL INFORMATION:

APPLICANT: Weinschank, Richard L.
APPLICANT: Branchek, Theresa
APPLICANT: Hartig, Paul R.
TITLE OF INVENTION: DNA ENCODING A HUMAN 5-HT1F RECEPTOR AND
NUMBER OF INVENTION: USES THEREOF
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: Cooper & Dunham
STREET: 30 Rockefeller Plaza
CITY: New York.
STATE: New York.
COUNTRY: USA
ZIP: 10112

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/817,920
FILING DATE: 19920108

CLASSIFICATION: 514

ATTORNEY/AGENT INFORMATION:

NAME: White, John P.
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 1795/39318
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-977-9550
TELEFAX: 212-664-0525

INFORMATION FOR SEQ ID NO: 8:

SEQUENCE CHARACTERISTICS:
LENGTH: 471 amino acids
TYPE: AMINO ACID
STRANDEDNESS: unknown
TOPOLOGY: Linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: N-terminal
IMMEDIATE SOURCE:
CLONE: 5-HT2
US-07-817-920-8

Query Match 5.5%; Score 91.5; DB 1; Length 471;
Best Local Similarly 20.4%; Pred. No. 0.1;
Matches 69; Conservative 52; Mismatches 77; Indels 141; Gaps 21;

Qy 45 LTIMLSMLFLVPFQSSLSSTSH---ISFKQAFGLGIIP-----CY----- 85
Db 94 LVIMAVS-----LEKRLQATNTYFLMSLAIDLLGLFVAPVSMLLTYGYRPLDSK 146
Qy 86 -----YALPNPQFQSF---VDLTKCVAMLLTYCG 115
Db 147 LCAWIVLYDLVLFSTASIMHLCALISDRVYALONPIHHSRFSNRTKAFKIIA-VWTISVG 205
Qy 116 PVLDVLYHLNPK-----SILEDYHEFLINTSFNRFIAPRTTEIFTSMLLTYL 169
Db 206 ISMPIPVGLDDSKVKEGSCILAD--DNEVLIGSFVSF-FIPL-----TIMVITYFL 256
Qy 170 NL-----IPHSOLSYOOLFW-----QPSLFGLAHAHAY 199
Db 257 TIKSLQKATLCVSDLGTRAKLASFLPQSSLSSEKLFORSIHREPGSYTG----- 308
Qy 200 EQLQSGMTVYSILTTGFOILYTTLFGGLTKFVFRGTGNLMC-----CITLHALC 251
Db 309 ----RRTMOSISNEQKAC-KYL-----GIVFELFV-----VWMCFFITINIMAVICKRSC 353
Qy 252 N--IMGPPGSRNLHFTVVDKKGRIKLY-SIWNKCY 287
Db 354 NEDVIG----ALLNV-FWIGYLSAVNPPLYTYTFENKTY 387

RESULT 7

US-08-370-542-7
Sequence 7, Application US/08370542
Patent No. 5476782

GENERAL INFORMATION:

APPLICANT: Weinschank, Richard L.
APPLICANT: Branchek, Theresa
APPLICANT: Hartig, Paul R.
TITLE OF INVENTION: DNA ENCODING A HUMAN 5-HT1E RECEPTOR AND
NUMBER OF INVENTION: USES THEREOF
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Cooper & Dunham
STREET: 30 Rockefeller Plaza
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10112

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/370,542
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/194,113
FILING DATE:
APPLICATION NUMBER: US/07/803,626

ATTORNEY/AGENT INFORMATION:

NAME: White, John P.
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 1795/39317
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-977-9550
TELEFAX: 212-664-0525

INFORMATION FOR SEQ ID NO: 7:

SEQUENCE CHARACTERISTICS:
LENGTH: 471 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: protein

TELEPHONE: 212-977-9550
 TELEFAX: 212-664-0525
 TELEX: 422523 COOP UI
 INFORMATION FOR SEQ ID NO: 8:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 471 amino acids
 TYPE: amino acid
 STRANDEDNESS: unknown
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 HYPOTHETICAL: NO
 ANTI-SENSE: NO
 FRAGMENT TYPE: N-terminal
 IMMEDIATE SOURCE:
 CLONE: 5-HT2
 US-08-216-594-8

Query Match 5.5%; Score 91.5; DB 1; Length 471;
 Best Local Similarity 20.4%; Pred. No. 0.1;
 Matches 69; Conservative 52; Mismatches 77; Indels 141; Gaps 21;

45 LTIMLISNLEFLVPLQSLSTSH---ISFKDAFLGLIIP-----GY----- 85
 94 LVIMAVS-----LEKKLQATNVPFLMSLAIDMLGLVMPVSMILTYGYWPLPSK 146
 86 -----YALPNWQSFQ---VKDLTKCYAMLLTYCG 115
 147 LCAWVIYLDVLFSTASIMHLCASLDRVVAIONPIHHSRFRSKRAKLIKIA-VWITSVG 205
 116 PVLDVLYHLNPK-----SSLEDYHEFLNWSFRNFIFARITEIFYSMLTTYL 169
 206 ISMPFIPFGQLDDSKVKESGCLAD--DNFVLIGSVSF-FIPL-----TWIVTYFL 256
 170 NL-----IPHSLSYQQLFW-----QPSLFEGLAHAAHAY 199
 257 TIKSLQKATLCVSDLGTRAKLASFSFLPQSSLSSEKLFORSIHREGSYG----- 308
 200 EQLOEGSMITVYSLITTCFOILYTLTFLGGLTKFVFVTRGMLWC-----CIILNALC 251
 309 ---RRTMOSISNQKAC-KVL-----GIVFELFV---VMCPFFITINIMAVICKESC 353
 252 N-IMGFPGPSRLNLFHTVYDKKAGRISKLV-SIWNKCY 287
 354 NEDVIG---ALLNV-FWVIGYLSAVNPVLYTLENKTY 387

RESULT 10
 US-08-542-358-7
 Sequence 7, Application US/08542358
 Patent No. 5786155
 GENERAL INFORMATION:
 APPLICANT: Weinschenk, Richard L.
 APPLICANT: Branchek, Theresa
 APPLICANT: Hartig, Paul R.
 TITLE OF INVENTION: DNA ENCODING A HUMAN 5-HT1E RECEPTOR AND USES THEREOF
 NUMBER OF SEQUENCES: 12
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Cooper & Dunham LLP
 STREET: 1185 Avenue of the Americas
 CITY: New York
 STATE: New York
 COUNTRY: U.S.A.
 ZIP: 10036
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/542.358
 FILING DATE:
 CLASSIFICATION: 435.

ATTORNEY/AGENT INFORMATION:
 NAME: White, John P.
 REGISTRATION NUMBER: 28,678
 REFERENCE/DOCKET NUMBER: 1795/39317-22/JPM/MAT
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 212-278-0400
 TELEFAX: 212-391-0525
 INFORMATION FOR SEQ ID NO: 7:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 471 amino acids
 TYPE: amino acid
 STRANDEDNESS: unknown
 TOPOLOGY: unknown
 MOLECULE TYPE: protein
 HYPOTHETICAL: NO
 ANTI-SENSE: NO
 FRAGMENT TYPE: N-terminal
 US-08-542-358-7

Query Match 5.5%; Score 91.5; DB 1; Length 471;
 Best Local Similarity 20.4%; Pred. No. 0.1;
 Matches 69; Conservative 52; Mismatches 77; Indels 141; Gaps 21;

45 LTIMLISNLEFLVPLQSLSTSH---ISFKDAFLGLIIP-----GY----- 85
 94 LVIMAVS-----LEKKLQATNVPFLMSLAIDMLGLVMPVSMILTYGYWPLPSK 146
 86 -----YALPNWQSFQ---VKDLTKCYAMLLTYCG 115
 147 LCAWVIYLDVLFSTASIMHLCASLDRVVAIONPIHHSRFRSKRAKLIKIA-VWITSVG 205
 116 PVLDVLYHLNPK-----SSLEDYHEFLNWSFRNFIFARITEIFYSMLTTYL 169
 206 ISMPFIPFGQLDDSKVKESGCLAD--DNFVLIGSVSF-FIPL-----TWIVTYFL 256
 170 NL-----IPHSLSYQQLFW-----QPSLFEGLAHAAHAY 199
 257 TIKSLQKATLCVSDLGTRAKLASFSFLPQSSLSSEKLFORSIHREGSYG----- 308
 200 EQLOEGSMITVYSLITTCFOILYTLTFLGGLTKFVFVTRGMLWC-----CIILNALC 251
 309 ---RRTMOSISNQKAC-KVL-----GIVFELFV---VMCPFFITINIMAVICKESC 353
 252 N-IMGFPGPSRLNLFHTVYDKKAGRISKLV-SIWNKCY 287
 354 NEDVIG---ALLNV-FWVIGYLSAVNPVLYTLENKTY 387

RESULT 11
 US-08-244-434-2
 Sequence 2, Application US/08244434
 Patent No. 5834004
 GENERAL INFORMATION:
 APPLICANT: Czerullofsky, A.P.
 APPLICANT: Himmler, A.
 APPLICANT: Stralowa, C.
 APPLICANT: Weyer, U.
 APPLICANT: Lameche, H.
 APPLICANT: Sch fer, R.
 TITLE OF INVENTION: Process for Screening Substances Capable of
 Modulating a Receptor-Dependent Cellular Signal
 TITLE OF INVENTION: Transmission Path
 NUMBER OF SEQUENCES: 37
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Sterne, Kessler, Goldstein & Fox
 STREET: 1100 New York Avenue, Suite 600
 CITY: Washington
 STATE: D.C.
 COUNTRY: U.S.A.
 ZIP: 20005
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/244,434
FILING DATE: 25-MAY-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/92/02718
FILING DATE: PCT Filing Date: 25-NOV-1992
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Esmond, Robert W.
REGISTRATION NUMBER: 32,893
REFERENCE/DOCKET NUMBER: 0652.1340000/RWE
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 371-2600
TELEFAX: (202) 371-2540
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 471 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-244-434-2

Query Match 5.5%; Score 91.5; DB 2; Length 471;
Best Local Similarity 20.4%; Pred. No. 0.1;
Matches 69; Conservative 52; Mismatches 77; Indels 141; Gaps 21;

QY 45 LTIMLISNLFVPLQSLSTSH---ISFKAPFLGLGIIP-----GY----- 85
DB 94 LVIMAVS-----LEKRLQATNYFLMSLAIDMLGLFVMPVSMTLTYGYWPLPSK 146
QY 86 -----YAAPNPWFQSF---VKDLTKCVAMLLTYCG 115
DB 147 LCAWITLDVLFSTASIMHLCALSIDRYVAIQNIHHSRFRSRKAKLKITA-VWITSVG 205
QY 116 PVLDVLYHLNPK-----SILEDIFYHEFLINWISFRNFIFADITEIEFTYMLTTYL 169
DB 206 ISMPIPVFGLQDDSKVEKESGLAD--DNFVLIGSFVSF-FIPL-----TIWVITYFL 256
QY 170 NL-----IPHSQSLYQQLFW-----QPSLFFGLAHAAHAY 199
DB 257 TIKSLQEARLQVSDICTRAKLASFSLPOSSLSEKLFORSIHREGSYTG----- 308
QY 200 EQLOEGSMITVSIILTTGFOILYTLTFLGGLTKFEVFRGTGNLMC-----CIILHALC 251
DB 309 -----RRTMOSISNQKAC-KYL-----GIYFLFLV---VMGCPFITINIMAVICKESC 353
QY 252 N-IMGFPGSRNLHFTYVDKKGARISKLV-SIWNKCY 287
DB 354 NEDVIG---ALLNV-FVMIGYLISSAVNPVLYTLEFNKTY 387

RESULT 12
US-09-018-351-7
Sequence 7, Application US/09018351
Patent No. 6096507
GENERAL INFORMATION:
APPLICANT: Weinshank, Richard L.
APPLICANT: Branchek, Theresa
APPLICANT: Hartig, Paul R.
TITLE OF INVENTION: DNA ENCODING A HUMAN 5-HT1E RECEPTOR
TITLE OF INVENTION: AND USES THEREOF
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Cooper & Dunham LLP
STREET: 1185 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.

ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/018,351
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/542,358
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: White, John P.
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 1795/39317-22/JPM/MAT
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-278-0400
TELEFAX: 212-391-0525
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 471 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: protein
HYPOTHEICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: N-terminal
US-09-018-351-7

Query Match 5.5%; Score 91.5; DB 3; Length 471;
Best Local Similarity 20.4%; Pred. No. 0.1;
Matches 69; Conservative 52; Mismatches 77; Indels 141; Gaps 21;

QY 45 LTIMLISNLFVPLQSLSTSH---ISFKAPFLGLGIIP-----GY----- 85
DB 94 LVIMAVS-----LEKRLQATNYFLMSLAIDMLGLFVMPVSMTLTYGYWPLPSK 146
QY 86 -----YAAPNPWFQSF---VKDLTKCVAMLLTYCG 115
DB 147 LCAWITLDVLFSTASIMHLCALSIDRYVAIQNIHHSRFRSRKAKLKITA-VWITSVG 205
QY 116 PVLDVLYHLNPK-----SILEDIFYHEFLINWISFRNFIFADITEIEFTYMLTTYL 169
DB 206 ISMPIPVFGLQDDSKVEKESGLAD--DNFVLIGSFVSF-FIPL-----TIWVITYFL 256
QY 170 NL-----IPHSQSLYQQLFW-----QPSLFFGLAHAAHAY 199
DB 257 TIKSLQEARLQVSDICTRAKLASFSLPOSSLSEKLFORSIHREGSYTG----- 308
QY 200 EQLOEGSMITVSIILTTGFOILYTLTFLGGLTKFEVFRGTGNLMC-----CIILHALC 251
DB 309 -----RRTMOSISNQKAC-KYL-----GIYFLFLV---VMGCPFITINIMAVICKESC 353
QY 252 N-IMGFPGSRNLHFTYVDKKGARISKLV-SIWNKCY 287
DB 354 NEDVIG---ALLNV-FVMIGYLISSAVNPVLYTLEFNKTY 387

RESULT 13
US-09-032-742-8
Sequence 8, Application US/09032742
Patent No. 6255089
GENERAL INFORMATION:
APPLICANT: Teitel, Milt
APPLICANT: Herrick-Davis, Katharine
APPLICANT: Egan, Christina C.
TITLE OF INVENTION: Constitutively Activated Serotonin
TITLE OF INVENTION: Receptors
NUMBER OF SEQUENCES: 25

| | | | | | | | |
|-----------------------|-------|--------------|------|------------|----|--------|-----|
| Query Match | 5.5% | Score | 91.5 | DB | 4 | Length | 471 |
| Best Local Similarity | 20.4% | Pred. | No | 0.1 | | | |
| Matches | 69 | Conservative | 52 | Mismatches | 77 | Indels | 141 |
| | | | | | | Gaps | 21 |

```

QY      45 LTIMLISMLFLVPELOSISTTSH-----ISFKDAFLGLTIIP-----Gy-----85
Dd      94 LVMAVVS-----LEKKLOQATVYFLMSLAIDMLLGLFLVMPYMSLTIIXGYWPLPSK 146
QY      86 -----YALPNWQFSQF---VKDLTKVAMLLLYCG 115
Dd      147 LCAWVIYLDVLEFSPASIMHLCAISLDLRVVALQNPINHSRFSRTRKAKELKIIA-VMTISVG 205
QY      116 PLVDFVLVHLNPK-----SLIEDFYHEFLNTWMSRNFLEAFITEIEFYTSMLLTYL 169
Dd      206 ISMPVPGLODDSKYKREGSCLAD--DNFVLIGSVSF-FIFL-----TIMVTIYFL 256
QY      170 NL-----IPHSOLSYQDLFW-----QPSLEFGLAHANHAAY 199
Dd      257 TIKSLQKEATPLCVSDLGRTRAKLASFSFELPOSSLSEKELFGRSIIHREGSYTG-----308
QY      200 EQLOEGSMATVYSILITTCFOFLITVTLFGGLKRFYFVGTGMLNC-----CIILHALC 251
Dd      309 -----RRMIOGISNSQKAC-KYL-----GIYFLFLV-----VMCPFFITIMAVICESC 353
QY      252 N-IMGPPGSRNLNHTFYVDKKGAKRISKLV-SIWNKCY 287
Dd      354 NEDVIG-----ALLNV-FWIGIGYLISSAVNPVLYTTFENKTY 387

RESULT 14
PCT-US93-00149-8
Sequence 8, Application PC/TUS9300149
GENERAL INFORMATION:
APPLICANT: Welnsheank, Richard L.
APPLICANT: Branchek, Theresa
APPLICANT: Hartig, Paul R.
TITLE OF INVENTION: DNA ENCODING A HUMAN 5-HT1F RECEPTOR AND
TITLE OF INVENTION: USES THEREOF
NUMBER OF SEQUENCES: 9

```

| | | | | |
|-----------------------|-----------------|---------------|------------|------------|
| Query Match | 5.5% | Score 91.5 | DB 5 | Length 471 |
| Best Local Similarity | 20.4% | Pred. No. 0.1 | | |
| Matches 69 | Conservative 52 | Mismatches 77 | Indels 141 | Gaps 21 |

| | | | | | |
|----|-----|--|--|---------|----|
| QY | 45 | LTLMHLSMLFLVPLPQLQSLSTSTSH---- | ISFKNAFLGLGIIP----- | Gy----- | 85 |
| | | : : | : : : : : | : : : | |
| Db | 94 | LVMMAVS----- | LEKKLQNAIYPLMMSLIAIOMLGLFVMPMSMLTIDYGYRMLPSK | 146 | |
| QY | 86 | ----- | YAAPNPWQESOF-----VKDITKCVAMLLTYCG | 115 | |
| | | : : | : : : : : | : : : | |
| Db | 147 | LCAWITLYLDFSTASIMHLCAISLDIRVAIQNIHHSRENSRKAKLKIIA-VWITISVG | 205 | | |
| QY | 116 | PVLDFVLVHLNPK----- | SLIDEFYHEFLNIMWSFNFIFABITEIEFYVTSMLTTYL | 165 | |
| | | : : | : : : : : | : : : | |
| Db | 206 | ISMPLPVPGLDDQSKVFKESGLIAD--DNVFLIGSVSF-FIPL----- | TIVITTYFL | 256 | |
| QY | 170 | NL----- | IPHSLSYQQLFW-----QPSLFFGLAHAAHAY | 199 | |
| | | : : | : : : : : | : : : | |
| Db | 257 | TIKSLOKEATLCVSDLGTRAKLASFSFLPOSSISEKLFORSIHREBSYTG----- | 308 | | |
| QY | 200 | EQLDEGSMVTYSILLTTCFOLYLTTLLEGGLTKPFVATGNCMLC----- | CIILHALC | 251 | |
| | | : : | : : : : : | : : : | |
| Db | 309 | ----RRTMOSISNEOKAC-KVL----- | GIFFLELV-----VMMCPFEITIMAVICKESC | 353 | |
| QY | 252 | N--IMGFPGPSRLNLHFTVVDKRGARISKLV-SIWNKCY | 287 | | |
| | | : : | : : : : : | : : : | |
| Db | 354 | NEDYIG-----ALLNV-FVMIGIYSSAIVNPLVYITLENKTY | 387 | | |

RESULT 15
 US-07-996-772A-11
 ; Sequence 11, Application US/07996772A
 Patent No. 547286
 GENERAL INFORMATION:

THIS PAGE BLANK (JPT0)